Partnerships for Evaluating Standards-based Professional Development for Teachers

Jacqueline Raphael
Michael Puma
(with assistance from Shannon McKay and Kristen Olson)
The Urban Institute

Chapters written by staff and evaluators with the following school districts:

Auburn School Department (Maine)
Jasper City School System (Alabama)
Oakland Unified School District (California)
The Ohio Valley Educational Cooperative (Kentucky)
Plainfield Public Schools (New Jersey)

2001

Submitted to:

U.S. Department of Education
Planning and Evaluation Service
400 Maryland Avenue, SW
Washington, DC 20202

Submitted by:

The Urban Institute
Education Policy Center
2100 M Street, NW, Ste. 500
Washington, DC 20037

Contract No.: EA94053001

TABLE OF CONTENTS

Partnerships for Evaluating Standards-based Professional Development for Teachers: A Summary

Jacqueline Raphael
Michael Puma

Systemic, standards-based reform emphasizes the need for school districts to be accountable to the public for their programs. With state and district academic standards commonly in place, districts' staffs must ask whether their teachers are effectively prepared to teach students to meet high challenging standards and whether district efforts to improve classroom instruction through professional development for teachers are contributing to increases in student academic achievement.

To answer such questions, districts need the capacity to monitor and evaluate their reform activities and to use data to inform ongoing systemic changes. It is easy to recognize the potential of a particular professional development program for teachers — even to select or develop one that appears likely to meet the district's needs and improve instruction. But to *measure* a professional development program's impact in terms of student achievement gains is quite difficult.

This report describes a project designed to help school districts address this challenge and illustrates the successes and challenges five partner districts encountered in evaluating the impact of key teacher professional development programs.

The final reports, which follow this summary, were written by district staff and/or outside evaluation consultants and based on evaluations that they conducted, with guidance from The Urban Institute.

Background

The Partnerships for Evaluating Standards-based Professional Development for Teachers Project was designed to combine the best of local initiative with technical expertise for the purpose of studying how to measure the impact of programs for helping teachers improve classroom instruction in a standards-based environment. Five partnerships were established with districts that: (1) were implementing professional development programs aligned with state or district standards and student assessments; (2) showed a desire to design and implement an evaluation of their professional development program; and (3) had access to two to three years of student achievement data that could be used in the evaluation. District partners agreed to share information about and incorporate feedback on the conduct of their evaluations; in turn, they received support and technical assistance for carrying out their partnership evaluation projects. This report describes the five partnership projects, their results, and lessons learned individually and collectively

from the projects.

The Partnership Districts

The Partnership Project involved a total of eight school districts — four single districts and a consortium of four small rural districts. The district partners represent urban and rural districts from regions across the United States. The professional development programs for teachers evaluated by the partners also represent a wide range of approaches to training teachers to use standards to improve instruction. In addition to the variation in the focus of their professional development activities, the district partners represent a variety of approaches in how they chose to implement and evaluate their programs.

- ***Auburn School District***, a small school district in a mid-sized city in Maine, focused its partnership project on the impact of the implementation of the Literacy Collaborative, a partnership among several Maine school districts, the University of Maine at Orono, the Maine State Department of Education, and Ohio University to improve literacy instruction among K-2 staff. The professional development program was systemic, focusing on classroom pedagogy, school leadership, teacher professional development, and student assessment. Lead teachers trained other teachers, including special education, Title I, and Reading Recovery teachers.

  For the partnership project, Auburn explored relationships between teachers' level of implementation and student achievement in reading and writing, as well as the relationship between teacher experience with the professional development program and the depth of implementation of the training that they achieved. Classroom teachers self-reported their depth of implementation, with their literacy coordinators providing a second measurement of their implementation level. In addition, the district analyzed student achievement data from three school years, across approximately 70 classrooms.

- ***Jasper City Schools***, a small rural district in Alabama, studied its professional development activities in educational technology through a professional development program called the "Technology Literate Teacher." The program incorporated nationally-developed education technology standards and supported the district's formal long-range technology training plan. The professional development involved one-on-one coaching and group workshops to help teachers gain basic computer literacy skills and to integrate those skills into curriculum development.

  For the partnership project, Jasper assessed whether individualized teacher instruction was more effective than group training, and whether the program was associated with teachers' increased use of technology for instruction, individualized professional development, and communication with parents and students. The staff also attempted to investigate the connection between teacher training and students' use of school technology. To assess the level of teacher implementation of the knowledge and skills gained from the professional

development, the district used teacher interviews, student and teacher surveys, a performance-based on-line computer skills testing software, and a review of teacher lesson plans. The district employed a statistician to determine the effect of the professional development program on student test scores in reading, particularly the scores of low-performing students.

- ***Oakland Unified School District***, a large high-poverty urban district in California, focused on professional development for middle-school teachers. The program, "Core Values for Technology Literacy," is a comprehensive training program, based on district English/language arts and social studies standards and using technology, is designed to improve classroom instruction in English/language arts and the social sciences. Core Values used several training methods, including an intensive summer institute, teacher in-service training, peer collaboration, coaching by "master" teachers, and in-person and on-line mentoring. Over 300 teachers have been served since the 1998-1999 school year.

  For its partnership project, Oakland concentrated on determining how students responded to instruction from teachers who had participated in the Core Values program. This partnership project built on previous evaluation findings.

- ***The Ohio Valley Educational Cooperative***, made up of four small rural districts in western Kentucky, implemented Project SMART, a program of professional development for K-12 math and science teachers. The training program, which started in October 1999, consisted of a variety of delivery methods, including 30-hour summer institutes for teachers and the use of the trainer-of-trainers model. Training focused on inquiry-based mathematics and science activities. The professional development was aligned with state and district content standards in participating districts, as well as related needs identified by participating schools as part of their state-mandated consolidated plans.

  The Ohio Valley Cooperative focused on assessing the impact of Project SMART on 48 K-12 math and science teachers' attitudes and skill levels, as well as changes in their students' achievement. Interviews, surveys, reviews of lesson plans and anecdotal records, and classroom observations were conducted.

- ***Plainfield Public Schools***, a high-poverty district in New Jersey, conducted a professional development program for all K-8 teachers focusing on language arts. A variety of professional development activities were incorporated into this program, but the partnership project focused on the use of lead teachers as literacy coaches to assist their colleagues.

  For the evaluation, data were collected on the role of the literacy coaches in different schools, and on teachers' level of implementation of the training they had received through surveys completed by literacy coaches and students. These data were augmented with available district data on the individual teachers. Student achievement on both a

performance assessment and a district-developed criterion-referenced assessment was analyzed using the Tennessee Value-Added assessment model.[1]

Working Together

Technical Assistance

Technical assistance provided to district partners was intended to complement unique local understanding of their schools with expert knowledge of the challenges of evaluating educational programs and methods for addressing them. As this report illustrates, many logistical and technical challenges exist to designing and conducting evaluations of professional development activities. Evaluation staff at The Urban Institute, the U.S. Department of Education's contractor for the Partnership Project, helped the district partners refine their initial evaluation plans, particularly with regard to common areas such as obtaining consistent student achievement data across several school years, defining clear research questions, and selecting appropriate data collection methods. This process of refining locally-developed evaluation plans involved an initial "kick-off" meeting with district partners and ongoing telephone and e-mail contact.

The Partnership Project presented districts with several challenges in the design and implementation of evaluations, including the following:

1. Separating out the effects of a specific professional development program when multiple reform activities are underway in a district (e.g., the introduction of new curricula, assessments, classroom materials, or technology).

2. Linking evaluation methods to nontraditional methods for delivering professional development, particularly teachers mentoring or teaching other teachers.

3. Moving from implementation of professional development to end outcomes (e.g., increased student achievement), which involves linking a series of implementation and intermediate outcomes together.

4. Needing to use a complex evaluation design due to the complexity of linking professional development to student outcomes, e.g., the design may require a combination of quantitative and qualitative methods, or an experimental design with control groups.

[1] This model, which measures the value added to student academic performance by teachers based on their students' test-score gains, was developed by William Sanders. It was adopted by the state of Tennessee in 1992 and is being used in every Tennessee school district. It is described in an article by Jeff Archer in the May 5, 1999 edition of *Education Week*.

5. Creating a "learning organization" throughout the district, so that the district uses data to make decisions and continuously improve. Districts attempting to do so sometimes face political challenges.

Not all of these challenges could be addressed by the district partners, as discussed in the next section of this report. Notably, these are challenges faced not only by the district partners, but also commonly encountered in most evaluations of educational programs.

Conducting the Evaluations

The Research Questions

The district partners generally focused their evaluations on two questions:

(1) To what extent did teachers implement the professional development training in their classrooms; and

(2) To what extent was this related to teacher and/or student knowledge, skills, and attitudes, and to student achievement as measured on district, state, or nationally-normed tests.

By linking teacher behavior and student achievement, the districts attempted to assess achievement of intermediate and end outcomes. Each partner's approach is summarized below:

- Auburn (Maine) first explored whether greater teacher experience with the professional development (i.e., years of experience in the program) was associated with a greater degree of implementation of the curriculum that was the subject of the professional development. Measuring the "depth" of implementation of the curriculum was a useful exercise for the district, with lead teachers discussing what implementation of each aspect of the curriculum would look like.

- Jasper (Alabama) implemented a number of evaluation activities to ascertain the effect of technology training on teacher practices, both inside and outside the classroom (e.g., communication with parents), as well as teacher skills and attitudes. In this case, implementation of the professional development model was not defined by a set of specific activities. The technology program was designed to enhance teachers' computer skills. Because the district provided training and equipment to all of its teachers, the entire faculty was included in the evaluation (though not all completed the surveys and assessments).

- Oakland (California) investigated how students responded to instruction from teachers with varying levels of "involvement" in the professional development program, measuring involvement through a combination of scales including the number of workshops attended

and frequency of use of resources connected with the professional development program.

- The Ohio Valley Educational Cooperative in Kentucky investigated the effects of intensive professional development on instruction for the teacher leaders, a subset of the K-12 faculty that received nearly 120 hours of professional development between 1999 and 2001. Like Jasper, the consortium looked at changes in teachers' knowledge, attitudes, and behaviors, as well as student achievement and other outcomes.

- Plainfield (New Jersey) used district records and a teacher questionnaire to ascertain the extent and quality of training received by teachers, and a student interview to ascertain the extent to which students participated in (and understood) specific features of the language arts curriculum encouraged by the professional development. The district then linked this implementation information to student test scores using a structural equation model.

A variety of specific research questions were investigated in these evaluations, but a strong feature of the project was the districts' attempts to explore a variety of outcomes that together describe a program logic.

Also impressive was the variety of methods the district partners employed to answer these research questions. The data collection methods used included teacher and student surveys, classroom observations, and the analysis of student achievement data. Most of the districts attempted to use a measure to confirm or triangulate teacher self-reports on classroom practices — an important effort at reducing self-reporting biases. For example, in addition to using teacher interviews and surveys on instruction, the Ohio Valley Educational Cooperative developed an observation instrument based on the state's teacher evaluation form and the new state technology requirements. Under the supervision of their external evaluator, the team piloted and used the instrument to conduct the observations. Auburn collected teacher self-reported data on the level of implementation of the professional development training, and also surveyed the district's literacy coordinators (or lead teachers) for a second assessment of the teachers' implementation level. Jasper (Alabama) reviewed lesson plans and teacher-developed portfolios as part of its assessment of teacher practice. Plainfield (New Jersey) had intended to survey both literacy coaches and regular classroom teachers to determine whether the coaching component of the professional development had been implemented as planned, but the response rate from coaches was too low to make the data useful. However, Plainfield used student interviews to gauge the level at which classes followed the components of the language arts professional development program.

Although the districts addressed key questions and used a variety of data collection methods, they were less successful at using an experimental design, with control groups. In most cases, this was due to the fact that all teachers at certain grade levels received the professional development.[2] This full-scale model is in keeping with the concept of systemic reform, whereby districts seek to improve instruction for all students, making lasting changes in teacher capacity throughout the district — not merely in selected schools or classrooms.

In addition, the district partners used varying levels of sophistication in their statistical controls. Two districts employed a multivariate analysis, while others used only minimal controls for teacher and student characteristics. Several of the district partners had difficulty incorporating their statistical analyses into their reports, often simply reporting data and covariance analyses without discussing the significance of these procedures.

Despite these capacity issues, all of the evaluation partners attempted to analyze student achievement data across two to three years. However, two district partners (Jasper and Oakland) were unable to incorporate this information into their evaluations. Jasper focused on the effects on teachers of the professional development program and explored the use of only one specific instructional software to see if it affected students' standardized test scores in reading. Oakland had intended to use district writing assessment results, which are said to be well-aligned with the professional development program. However, these data were available to schools only by school, and the extent of professional development received varied greatly by teacher, not school. The district sought to make comparisons among a select number of schools, but the small sample size limited the usefulness of these data.

During the course of the project, several were pessimistic about finding any statistically significant results, and provided several reasons for their skepticism. One valid concern was the relatively short time frame for this project. Even with several years of data available, they felt that it would take more time for professional development training to have an effect on student achievement. Another concern, also valid, was the lack of measures of student achievement that were well-aligned with the objectives of their staff development. Although the district partners were interested in student outcomes, some were not confident that the state or district assessment was the best tool for measuring the changes in instruction, either because the tests were less specific to the district's instructional goals or were still in development. Most districts — including Plainfield, Oakland, and Auburn — indicated that their district tests were more relevant to the goals of their professional development programs than statewide assessments or nationally normed tests. However, in several instances, district test score data were available only by school and/or student — not by teacher — making the use of these data quite challenging. One district partner, the Ohio Valley Educational Cooperative, attempted to link student data to particular teachers by searching through district records but was unsuccessful.

[2] Jasper (Alabama) had hoped to compare data on teachers' computer skills with data from another Alabama school district that had implemented a similar training program, although teachers there participated on a voluntary basis. However, the other school district relied on teacher self-assessments rather than computer skill tests results.

As a proxy, the district focused on data in schools where there was only one science or mathematics teacher at that grade level in the school.

Concerns about incorporating student test score data into the evaluations were discussed frequently during the project, including the possibility that none of the partners would identify statistically significant results in student test scores. Despite these limitations, the districts "pushed the envelope" by at least attempting to use student test score data in their evaluations. Furthermore, by conducting these evaluations and making their efforts public, they raised important issues for other districts who may want to conduct similar research.

Lessons Learned

As the partners individually explored whether their efforts to improve classroom instruction through standards-based professional development for teacher were contributing to increases in student academic achievements, several common themes emerged. Identified through the processes of planning, conducting, and reporting on the partner evaluations of professional development programs, these themes represent factors that can strengthen or weaken evaluations and fall into the four categories highlighted below.

Initial Design Constraints

Local program staff have different needs for evaluation information. District staff, particularly those trying to respond to the new demands for greater school accountability, often request student outcome data to show that the program is "working." On the other hand, professional development staff, teachers, and others closer to the "ground floor" may be more interested in other outcomes, such as the impact of staff development on teachers. Trying to accommodate various stakeholders' needs can be difficult, particularly when districts have limited funding for evaluation. One key to balancing stakeholders' needs is to establish a long-range evaluation plan that includes short-term projects along the way that can satisfy various stakeholders' needs. In addition, evaluators must stress that not all questions can be answered in a single evaluation.

Local program staff may be under pressure to produce positive results quickly. Professional development does not necessarily impact student achievement immediately. In fact, instructional quality can actually decrease when teachers initially attempt to use new techniques in their classrooms. Furthermore, the kind of fundamental instructional change standards-based reform may require can take teachers several years to accomplish. District staff may know this, but at the same time feel pressured to show that their programs are working, often within a year's time. For this reason, it is important for evaluators, practitioners, and district staff to be realistic with district staff about when anticipated program outcomes can reasonably be expected to occur.

Constraints on time and resources can limit the rigor of the evaluation. When school

districts can offer only minimal funding for an evaluation, yet need a quick turn-around for results, it can compromise the rigor of the evaluation. Evaluators therefore need to create a careful balance between trying to obtain valid answers to evaluation questions while remaining within the available resources of time and money. Significantly, three of the five district partners used an external evaluator who worked closely with local staff for this project, while one of the other districts used a statistician to assist with particular aspects of data collection or data analysis. Such supplemental services are not frequently budgeted for by schools districts.

It is particularly difficult for district evaluators to make the claim that a professional development program *caused* observed improvements when the district is engaged in systemic reform, which typically involves multiple reform activities occurring simultaneously. Districts can attempt to correct for this in several ways, one of which is to explore whether a plausible alternative explanation or cause for improvements occurred. The five district partners involved in this project did not have the resources required for this additional investigation. As a consequence, several opted to explore whether more training was associated with greater change/effect. The one-year time constraint for this study probably contributed to lack of statically significant findings as well.

Finally, as one district partner pointed out, districts sometimes neglect to focus sufficiently on the reliability and validity of their data collection instruments — or may do so only until the need to use the instruments for an evaluation is pressing. The district may then feel forced to use measures with low internal consistency and reliability.

Constraints due to the nature of the treatment imposed limits on the evaluation design.

The use of a randomized or experimental design — in which participants (e.g., teachers) are randomly assigned to receive either the treatment (e.g., the professional development program) or to receive no services (referred to as the control or comparison group) — is the "gold standard" in evaluation. Yet few districts can easily randomly assign teachers to a professional development programs — and many districts choose to offer new professional development initially to highly motivated teacher volunteers. If the program is deemed successful, the district may offer the program to greater numbers of teachers in the future (or, may use these teachers to train others in the district). By not randomly assigning teachers to treatment and non-treatment groups, districts will find it nearly impossible to identify and use a comparison group in an evaluation.

This leads to another issue: ultimately, *systemic* professional development is provided to a large portion, if not all, of the teaching force in the district. That is part of what makes it *systemic* — it saturates the system. For this reason, then, most of the five partners — which had already been engaged in systemic reform for some time — did not use comparison groups. Instead, some attempted to compare outcome data to standards such as nationally standardized test score data and/or to outcomes achieved by other districts. Small districts (e.g., 75 teachers total) face an additional design constraint: the size of their teaching force makes it difficult to find statistically significant variation in their results.

Understanding the Program

Engaging practitioners in the design of an evaluation can help the evaluators better understand the program. Engaging local program staff in an evaluation begins with a dialogue about the expected outcomes of the program. Consequently, many of the district partners worked with the designers of the respective professional development activities to ensure that an understanding of the program goals and implementation history was built into the evaluation design.

Another important part of the evaluation teams' work was to specify the underlying logic of the program to support the design of an evaluation. This program logic explains how observed effects might be attributed to the program by describing the sequential steps that were expected to lead to the desired end outcomes (e.g., increased student achievement). District partners identified important "intermediate outcomes" such as the extent of participants' attendance at professional development sessions, changes in attitudes and behaviors due to the professional development, and actual changes in classroom pedagogy. If the expected intermediate effects were not observed, then it was difficult to argue that observed changes in student outcomes could have been the result of the program intervention.

Gaining Access to Information

Getting access to local program staff can be a problem. Those conducting evaluations may find it difficult to speak to school and district staff during business hours. Teachers, in particular, are often not accessible by telephone, and district staff and professional developers may also be hard to reach. Coordinating meetings between staff from different schools or district offices can also be challenging.

Using administrative records can also be difficult. Another challenge related to evaluation in school districts is that these institutions are not always organized to facilitate the use of data for evaluation. In one case, information on student attendance and grades, and information on teachers, were kept in different databases. In addition, teacher training data were not structured to readily provide the number of hours that each teacher spent on professional development. Inaccurate and incomplete data also complicated most of the five evaluations.

Using student test score data may also be difficult for evaluators. Two district partners found that student test score data — in one case, state tests, in the other, district tests — was reported to districts by school and student, but not by teacher. This either prevented using the test score results, or required that the evaluation team "hand link" teacher and student records to make the necessary connections. In this example, the evaluation team also had to collect these data while maintaining the confidentiality of both teachers and students.

Student test score data may be politicized. Although state, and some district, accountability systems rely on student assessment data, the use of such data is sensitive and highly politicized,

according to most of our district partners. Particularly in urban districts with chronically low test scores, staff can be very uncomfortable about the use of student assessment data in evaluations. Staff may be nervous about lowering school staff morale, often already suffering because of state or district accountability systems that "rate" schools publicly. Student assessment data are "loaded," we were told by our partners, and staff may resist dealing with analyses that could be interpreted as negative.

The Need for Capacity Building

Building local capacity is important but difficult. Each of the five district partners attempted to encourage the regular and effective use of data by involving a broad cross section of staff in a continuous improvement process that increased collaboration and the use of information to guide program development and growth. Although this process can, if implemented well, strengthen the organization, it can often seem like an impediment to an evaluator who is trying to conduct a technically sound study in a timely manner.

Practitioners have a limited understanding of evaluation concepts. Some of the districts reported problems associated with the limited understanding of program evaluation on the part of local practitioners — particularly data collection and the ~~concepts of~~ internal and external validity, reliability, and generalizability of data. This lack of understanding could, in their view, compromise the quality and/or usefulness of their evaluations. We collect "plenty" of data, said one of the district contacts, a staff member with an unusual degree of evaluation experience, "and I have control over most of it. But, I cannot convince our administrators of the sorts of things that need to be examined to ensure that it is quality data. Hence, we have the age-old problem of only being able to get the level of quality out of a body of data that you are willing to put in." Again, this points to a need for an ongoing dialogue between evaluators and practitioners. To address the lack of understanding of evaluation in the district described above, this individual plans to work with a local evaluation specialist to train program staff on evaluation.

Maintaining continuity of staff is also an issue. Staff turnover is a serious problem for a long-term evaluation. During the course of the partnership projects, two of the five district contacts left their positions. These individuals were advocates for the effective use of evaluation in their districts, and it is likely that they served as district leaders in other areas as well. It is not surprising, then, that they were chosen (or themselves chose) to tackle new reform challenges. Those on the front lines of district change are probably the most likely to change jobs, presenting a challenge for any ongoing evaluation.

Those involved in delivering and evaluating the professional development may also leave a district, posing additional challenges to both program delivery and program evaluation. For example, the smallest district partner lost both of its staff trainers, as well as the evaluation statistician. This district's training and data analysis had to be delayed, and valuable time was lost interviewing candidates to replace the lost personnel. In small districts, staff is often

stretched thin. Staff turnover can complicate the implementation of an evaluation, as can the loss of an external evaluator.

Collaboration and communication is needed throughout the evaluation process. For many reasons, school and district staff may feel threatened by evaluation. One district contact told us that some teachers in her district were "wary" of evaluation, even "uncomfortable" and "anxious" about it, in part because they were not accustomed to close scrutiny of their instructional practice. District staff may also feel uncomfortable about evaluation. On the other hand, evaluators (or those with evaluation experience) may feel uncomfortable when challenged by district and school staff about their study. These concerns can surface during the design, data collection, data analysis, and reporting stages. The best way to avoid creating a "we-them" dynamic in the evaluation process is to conduct the evaluation as a team, working collaboratively with a representative teacher, district staff member, professional developer, or others.

Evaluators need to communicate regularly throughout the evaluation with teachers and other staff who will provide data. Teachers are more likely to complete surveys and participate in other evaluation activities if: (1) they are informed early on about the purpose of the evaluation, and are well-informed in advance about evaluation activities; (2) if methods for ensuring confidentiality are explained to them; (3) if they are kept abreast of the progress of the evaluation; and (4) if evaluators share their results with teachers, preferably in a report or presentation tailored to this audience's interests.

One outside evaluator working with one of the five districts found that the best way to get data from schools and teachers was to find someone on campus willing to participate in the evaluation (i.e., talk to teachers face to face about the evaluation procedures and engage their support to facilitate data collection, particularly with non-respondents). In this way the evaluation can became a true presence at the participating schools instead of an occasional bother.

Conclusion

The five district partners involved in the Partnership Project conducted more comprehensive or in-depth evaluations than they typically have been able to do. In several cases, the district used the results of these evaluations to "stay the course" in providing standards-based professional development for teachers. Even where sizable positive results were not realized, districts looked at ways to continue to conduct evaluations in order to see if, over a longer time period, more significant gains could be shown.

Most of the district partners were able to document positive changes in teachers' skills, attitudes, and/or classroom practices, but less able to demonstrate that these changes led to changes in student outcomes, particularly student achievement. As discussed earlier, this may be due to a lack of alignment between the tests used and the objectives of the professional

development programs. In some cases, the lack of use of several years' worth of test score data contributed to the difficulty in showing change.

Perhaps more significantly, district partners appreciated the opportunity to conduct more thorough evaluations that could potentially enhance their capacity for continuous improvement. As one of our partners indicated, smaller districts, unlike large urban districts, do not have research and evaluation departments. Typically, they do not have the resources required to undertake this type of evaluation. The funds provided by this project, though limited, were used to provide additional time and in three cases additional expertise, in the form of an external evaluator or statistician, to these evaluations.

Several lessons were learned about district evaluation through this project. These include the existence of substantial initial evaluation design constraints; the need to bring practitioners and evaluators together to thoroughly understand the program being evaluated and use the results to improve the program; the challenge of gaining access to needed data for evaluation; and the need to focus on capacity-building. Each of these lessons is discussed in specific terms in the evaluation reports that follow.

However, what is as yet unclear is whether the lessons learned through the design and implementation of these evaluations will lead to a future commitment to evaluation from these districts. None of the district partners appear to have developed a commitment in their district to continued evaluation. One more thorough evaluation will not change the capacity of these districts to use evaluation results in designing or conducting professional development for teachers. As we have seen, even energizing individual teams to do such an evaluation doesn't always lead to permanent change, as these individuals are not a permanent fixture in the district infrastructure.

What does this project suggest for policymakers interested in encouraging district use of professional development evaluation? Several issues are relevant:

1. District staff need support to conduct more rigorous evaluations. The districts participating in this project received additional funding, and technical assistance — and even these districts had difficulty designing and conducting their evaluations.

2. Access to local data can be a major obstacle to effective evaluation. State test score data are not always available by individual student, making it quite difficult to link teacher and student outcomes. In some cases the data had appeared likely to be available in a variety of formats but turned out not to be. It is important for continuous improvement that districts are clear about what format and type of data will be available on a regular basis so that evaluations can be better planned.

3. Collaboration among key stakeholders — including district and school staff, evaluators, and program developers — is not only useful to improve the quality of an evaluation, but is

likely to improve the chances that the evaluation results will be used by local policymakers. The district partners reminded us repeatedly of the value of collaboration. They stressed the need for an evaluation *process* that the district could internalize. This process would help them communicate to administrative and instructional staff throughout the district:

- the goals of the program to be evaluated;
- the time frame for achieving those goals;
- the advantages and disadvantages of various research and data collection strategies;
- how to interpret evaluation results appropriately, including the limitations of the data that are ultimately collected; and
- how to use program development to improve ongoing programs and plan for the future.

To help districts understand the effects of their professional development programs, evaluators need to adapt their methodological positions to the developmental stage in which the district is currently operating. Likewise, policymakers must consider the district's capacity to conduct evaluations as set against the need to deliver more services to increasingly more diverse students and teachers.

DISTRICT EVALUATION REPORTS

The final reports that follow were written by district staff and/or outside evaluation consultants and are based on evaluations that they conducted, with guidance from The Urban Institute.

LITERACY COLLABORATIVE
AUBURN SCHOOL DEPARTMENT
Molly Schen
Michelle Porche (Center for Research on Women, Wellesley College)
Also with Li Gowell, Pat Kordalski, Pam Ouimette, Zoanne Paradis, and Sue Card

This study explores the effects of a school district's effort to increase the literacy of young schoolchildren through intensive professional development of all kindergarten, first- and second-grade teachers. The research questions concern the relationship between teachers' exposure to the professional development and their implementation of the core components in the classroom; and the relationship between teachers' implementation and their students' reading and writing achievement. To answer these questions, the research team used a teacher survey, classroom observations, and student achievement data. Key findings include a statistically significant correlation between teachers' implementation of the professional development framework and grade 1 reading scores, as well as significantly higher reading levels for students receiving free or reduced-price lunch who had a teacher trained in the program.

Introduction

Like many school districts, the Auburn School Department in Auburn, Maine, recognizes the importance of children gaining literacy skills early in their school careers. Scores of well-designed studies conclude that early literacy acquisition substantially increases the likelihood of academic success. Academic success, in turn, augurs well for a student's transition into citizenship that is productive and well-informed.[1] In short, early literacy is seen as a key leverage point for achieving the school district's mission of "working collaboratively so that all students learn and succeed in a changing world."

In the early 1990s, the superintendent created a task force to recommend strategies for early intervention when children were struggling to read. After reviewing the literature and meeting with reading specialists, the district decided to train several teachers in Reading Recovery. The idea was to work intensively with a small number of students who were struggling the most and, in addition, to create small literacy groups for Title I children. The response from teachers and students was enthusiastic. Teachers expressed a desire for many more students to be served by Reading Recovery. At the same time, many teachers and administrators recognized a need for a robust framework and training in literacy instruction in all classrooms.

Auburn School Department is Maine's fourth-largest district, serving 4,003 students with 628 staff in 12 facilities. The per-pupil operating cost at the elementary level for 1998-99 was $3737.66. Over one-third (36 percent) of students live at or below the poverty level as determined by free and reduced-price lunch qualifications. The student population is mainly homogeneous and white.

[1] Murnane, R.J., & Levy, F. (1996). Teaching the new basic skills: Principles for educating children to thrive in a changing economy. New York: Free Press.

District personnel began to search for a framework that would affect even larger numbers of students. The district's goal was to improve early literacy skills for children in grades K-2. Realizing that good instruction in the early grades is crucial, district personnel explored many programs and instructional frameworks, including the Literacy Collaborative framework. The Literacy Collaborative (LC) is the brainchild of a group of Ohio educators (primary classroom teachers, Reading Recovery teachers, and university personnel at Ohio State University) who reviewed research on best practices in teaching early literacy. They designed the LC framework to complement the Reading Recovery program. LC provides all children with high-quality instruction in reading and writing, with Reading Recovery as a safety net. In Maine, LC is a collaborative effort among several school districts, the University of Maine at Orono, the Maine State Department of Education and Ohio State University. It provides long-term professional development and systemic support for educators who are responsible for providing literacy instruction in kindergarten through second grade. The program aims to improve literacy acquisition for all K-2 students, rather than relying solely on a pull-out program for the lowest 20 percent of children in each grade. The project specifies annual achievement targets as defined by the Maine Educational Assessment: 65% of first graders will perform at the basic level or above on text reading level tests.

In 1996, Auburn decided to train literacy specialists in the Literacy Collaborative model while maintaining the Reading Recovery program. The initial commitment was to implement LC for five years, at which time the program would be evaluated, and a decision would be made about whether or not to continue. Three teachers were selected to receive year-long, intensive LC professional development to become trainers for the district. This training included weekly training, regular readings and assignments, and classroom observations to see the LC framework in operation in others' classes. At the same time, teachers implemented the framework in their own classrooms, were observed and videotaped, and received regular feedback. The training culminated in a weeklong classroom demonstration of the LC framework. Interested teachers in the district visited and were invited to sign up for training the following year. Thirty-five (35) K-2 teachers participated in the first year of district training (1997-1998), representing 47% of the total number of K-2 teachers. The LC coaches provided in-class modeling, coaching, and co-teaching assistance to this first cohort of teachers.

In the spring of 1998, Auburn made LC training a condition of employment for all K-2 teachers within two years of their hiring. If teachers did not comply, they would be transferred to a higher grade level. In addition, once teachers completed their initial training, they were required to attend four continuing contact sessions each year, held during staff development hours on Wednesday afternoons. Topics for these sessions were generated by teacher input and delved deeper into theory. Once LC training was mandatory, an additional twenty-three (23) teachers signed up for the course in 1998-1999, comprising 31% of the total K-2 teachers. In 1999-2000, the last of the veteran K-2 teachers were trained, as well as some new staff, for a total of seventeen (17) teachers, bringing the total teachers trained in LC to 96% by the end of the third year (see Table 1.3 for comparisons of LC-trained teacher pool to survey respondents). Staff training has included classroom teachers, Title I teachers, Reading Recovery teachers, special

education teachers, and an English-as-a-second-language (ESL) teacher. The LC coaches have also introduced the LC framework to elementary principals and school committee members.

Analysis from 1999-2000 data revealed that grade 3 students were still not performing as well as expected. For example, it was expected that aggregated reading comprehension scores for Auburn schools would be well above the 50% national norm, given the emphasis on early literacy. However, two schools did not even meet the national average. Auburn fourth-grade reading scores on the state test were 265 for the district compared to 295 for students across the state.

Research Questions

This study addresses the following research questions:

1) What is the relationship between teacher experience with LC and depth of implementation of the LC framework?
2) What is the relationship between a teacher's depth of implementation of LC and student achievement in reading and writing?
3) Is the K-2 LC an effective professional development model to improve student literacy?

Methods

Virtually every aspect of the study was accomplished as a collaborative research team. The team was comprised of the director of learning and teaching, four LC coaches, a statistician, and a consulting teacher. Typically, the team met together a half-dozen times in the spring, and another half-dozen times in the summer and fall for two to six hours each meeting. We discussed common issues as a whole group, and broke into small work teams to divide work, and reconvened to review progress. There were two exceptions to this general method of operating. Statistical analysis and final draft writing were done by the statistician and director of learning and teaching, respectively. For the rest of the time, the team gathered for work sessions ranging from two to eight hours. The Auburn School Department's management team (including the superintendent and Title I coordinator) were kept apprised of developments at every turn.

The following table summarizes Auburn's research questions and data collection methods:

Table 1.1
Research questions, data collected, and data sources

Research Question	Data Collected	Data Sources
1. What is the relationship between teacher experience with LC and depth of implementation of the LC framework?	Year that each teacher was trained in LC; teachers' self-assessment on depth of implementation; LC coaches' assessment of the teachers' depth of implementation	Records of training; survey instrument (used by both teachers and LC coaches) on depth of implementation
2. What is the relationship between a teacher's depth of implementation of LC and student achievement in reading and writing?	Composite scores on depth of implementation on different parts of the LC framework; all available district data on K, 1, 2, and 3 students' achievement in reading and writing	Survey instrument; Kindergarten portfolio; instructional text level K-2; text book level K-2; Jerry Johns Informal Reading Inventory 3rd grade; Auburn writing assessment 3rd grade; and Iowa Test of Basic Skills 3rd Grade
3: Is the K-2 LC an effective professional development model to improve student literacy?	Merging of teacher data with student achievement data; a decision-making path that included recommendations from the research team, district's management team, and school committee.	Depth of implementation variables regressed against student achievement outcomes, with additional control measures; research team, management team, and school committee

Data Collection

Several years of data were examined in order to describe a range of teacher experience with the LC framework and a larger set of student achievement scores.

Student Participants in Study. Records from three consecutive school years (1997-1998, 1998-1999, 1999-2000), for all students in kindergarten through second grade, were included in the study. In the final year's data set, third grade students were also included. Table 1.2 provides a breakdown of the number of students included by year and grade, showing between one and three years of outcome data per student cohort. For example, the first grade cohort of 1997-1998 is followed into second and third grade (as indicated by the diagonal).

Table 1.2
Number of student study participants by year and grade

	1997-1998	1998-1999	1999-2000
Kindergarten		279	228
First Grade	281	280	277
Second Grade	249	294	279
Third Grade			303

Teacher Participants in Study. Teacher rosters for the three school years (1997-1998, 1998-1999, 1999-2000) included 75 teachers who were eligible for LC training. Of those, 62 actually completed the teacher survey. Two teachers had left the district by the 1999-2000 school year, five teachers did not have regular classrooms in which they could implement the LC framework (instead providing Title I or Special Education services), three had not received LC training and therefore could not respond to survey questions about its implementation, and three did not return the survey despite the best efforts of the research team to obtain their cooperation.

Eighty percent of the teachers had a bachelor's degree, 18 percent had a master's degree, and one teacher had a post-master's degree, a Certificate of Advanced Study. Among the respondents, a broad range of teaching experience was represented, from first year teachers to veterans with up to 38 years of experience. Overall, this was an experienced staff, with 75% of teachers having been in the classroom eight years or more. There was only one male teacher among the teacher participants.

Construct Development

Three constructs helped the team measure LC's effectiveness: "teacher experience with LC," "depth of LC implementation," and "student achievement in reading and writing."

Teacher experience with LC. "Teacher experience" is a simple measure of the teacher's years of experience with LC. Some teachers in Auburn took the LC course in 1997-98, so during our study at the end of 2000, they had three years of experience with the framework. At the end of 2000, there were still a few teachers who had not yet taken the LC course. The range of experience with LC as shown below in Table 1.3 is from 0 to 3 years.

Table 1.3
K-2 teacher pool compared to survey respondents by years of LC experience.

	3 Years LC Experience (starting 1997-98)	**2 Years LC Experience (starting 1998-99)**	**1 Year LC Experience (starting 1999-00)**	**No LC Training**
Number of K-2 LC-Trained Teachers	35	23	16	3
Number of K-2 LC-Trained Teachers Responding to Survey	29	18	15	3[2]

Depth of LC implementation. The second construct is "depth of implementation of LC." This was by far the most challenging—and rewarding—construct to operationalize. Our efforts resulting in the development and use of a survey organized by the components of the LC framework. The LC coaches discussed each aspect of the framework in rich detail. For example, what does it look like when teachers first begin to implement "shared reading," as compared with an experienced and more skillful implementation of this portion of the framework? Fortunately, the LC developers at Ohio State University had already constructed a continuum for two aspects of the LC framework, interactive writing and guided reading. These were modified to fit the needs of the study. The research team also developed continua for other aspects of the LC framework, including read-alouds, shared reading, writer's workshop, and letter and word work. The LC coaches worked through several drafts of a survey instrument that described behaviors typical of "early implementers" and "skilled implementers" in order to capture teachers' depth of implementation. The discussions in themselves were enriching to the LC coaches, who were able to share specific teacher behaviors they recognized along a continuum of development.

Student achievement in reading and writing. At first this third construct seemed simple. Data needed to be collected on student achievement in reading and writing. Reading and writing assessments are part of the standard school evaluation of student progress. All analyses were conducted using code numbers for students' data to maintain confidentiality and to allow for change comparisons over time. Once work had already begun on the construct, there was the realization that in Auburn, different kinds of data are collected at different grade levels. Thus, data analysis had to be sensitive to grade-level specific data. Table 1.4 below lists all of the measures and how achievement scores were translated into statistics for analysis.

[2] In addition, in 1999-2000, data were collected from 12 third grade teachers with no LC training in order to obtain follow-up information.

Table 1.4
Auburn achievement measures used in this study

Achievement measure / Grades Administered	**Description**	**Translation for this study**
Kindergarten portfolio Kindergarten	Four subtests (standard in the district) including letter identification, word test, concepts about print, hearing and recording sounds and words.	Percentage of correct responses out of 135
Instructional text level (1997-1998; 1998-1999) Grades 1-3	Rating according to book level for instruction: Distinguished, Advanced, Basic, Novice	D-A-B-N translated to 4-3-2-1
Instructional text level (1999-2000) Grades 1-3	Same kind of rating, but with new nomenclature to match the new state proficiency levels: Exceeds the standard, Meets the standard, Partially meets the standard, Does not meet the standard	E-M-P-D translated to 4-3-2-1
Text book level (1997-1998; 1998-1999: Grades 1-2; 1999-2000: Grades K-2)	18 levels of book readings from A to R, plus 0 for students not reading at the lowest level. Levels determined by Ohio State based on characteristics within the text. (R is defined as being placed in 4th grade, with expanded comprehension in evidence.)	A-R translated to 1-18, plus "0"
Auburn writing assessment Grade 3	Standardized writing prompt and scored by two trained teachers, using Maine's writing rubric. Scores range from 2-12.	D-A-B-N translated to 4-3-2-1
Jerry Johns Informal Reading Inventory Grade 3	Students read to themselves and answer questions afterwards. Range of scores correspond to grade level (0-8)	0-8
Iowa Test of Basic Skills (ITBS) Grade 3	3rd graders' scores on ITBS reading comprehension section, given as a percentile compared to a nationally normed sample	0-99

Challenges

Depth of LC Implementation. Several issues related to the construction of the survey arose. First, although we agreed on behaviors that characterize LC implementation, it was difficult to describe them in a value-neutral manner on a survey instrument. The team struggled to make statements that were neither positive nor negative. Teachers had to select a statement and then decide whether the statement was "very much like me," somewhat like me," or "a little like me." Auburn wanted teachers to be honest in describing their depth of implementation, but in the end, teachers were readily able to identify which end of the continua represented the "deep implementation" response (even though "deep implementation" was randomly placed on the right and the left sides of the page). Overall, most teachers rated themselves quite high on the continua.

A second problem was determining reliability of the LC coaches' own views of teachers' depth of implementation. If the teachers rated themselves very favorably on the continua, it made sense that LC coaches might be able to describe them more objectively. But would the LC coaches' views of teachers conflict? A system needed to be established whereby several teachers would be observed by two LC coaches, and then correlation tests run on the LC coach judgements. However, the need for additional classroom observation excited some controversy in the spring of 2000. Most classroom teachers were comfortable having their "own" LC coach come into the classroom for observation, coaching, modeling, etc., but not as open to welcoming another LC coach into their classroom, especially at the frenetic end of the year. At one point quite a few "double" observations were set up. With the reality of teachers' discomfort, the numbers were reduced considerably, after receiving statistical assurances that there would be enough data for testing correlations between LC coaches.

LC coaches evaluated teachers for whom they provided training and had observed in practice throughout the year. They used the same rating sheet the teachers used (the survey instrument). In cases where coaches felt they did not have adequate information to complete ratings, they did additional observations. In order to test for reliability, ten percent of the teachers were randomly selected from the entire sample for second observations. For the second set of ratings, LC coaches went to classrooms and spent approximately 90 minutes observing teachers' implementation of the LC framework.

We considered weighting certain aspects of the framework more than others, based on their prominence at certain grade levels. In the end, this was not done. However, we tested how specific components of the LC framework related to reading and writing outcomes.

Teachers raised some concerns about the study. The research team understood that teachers were unaccustomed to close scrutiny of their classroom practice, especially for a well-designed statistical study. In order to provide information to teachers about the study, we held an informational meeting, open to all K-3 staff. Approximately twenty teachers attended, and while some people voiced good questions, there was discussion that other teachers, not present, were somewhat disgruntled by the scrutiny of individual

teachers' practices. To assuage these teachers' concerns, the LC coaches reassured the various individuals about individual and school anonymity in the study's findings — particularly, that although we needed to be able to link teachers' self-assessments with their students' achievement, findings would be aggregated in the final report. No teacher nor individual school would be identified.

The survey was piloted to a randomly selected group of four teachers. As a result of their feedback, several minor revisions were made. At continuing contact sessions in May 2000, the LC coaches handed out the surveys to all trained K-3 teachers and specialists. In order to gain a 100% completion rate, staff were given time to complete the surveys at these sessions.

By planning ahead, we were able to incorporate two of our three concerns of the study in the survey instrument, that is measuring depth of implementation from the teachers' self-reports as well as coordinators' assessments of teachers' implementation, but not weighting of the various aspects of the LC program. The cover sheet of the survey collected crucial data including teacher name and years of teaching experience. It also contained the name of the teacher's LC coach. Coaches compiled lists of dates when teachers were trained, thus enabling the collection of data on each teacher's years of LC experience. We maintained teacher confidentiality through a complex but effective system using ID numbers that were placed on a separate cover page of the survey.

The total number of K-2 teachers in the sample was 75. There were three literacy coordinators, each of whom worked with 20-30 teachers. One challenge in the evaluation was that even though Auburn is one of the largest districts in the state of Maine, and all K-2 teachers and students were included in the study, the primary unit of analysis was the teacher. Seventy-five teachers does not constitute a very large study, especially when the analyses are conducted by grade, further reducing sample size. It turned out to be hard to see a strong, statistically significant variation in such a small sample.

Student achievement in reading and writing. The collection of these data, spanning three years, should have been easy, but it was not. The data were generated at different times of the year, from various sources ranging from commercial testing companies to district-designed databases. Some of the data were retrieved with the assistance of the technology director, who asked technology assistants to collect disks and student identification lists from each elementary school. The data were also retrieved from file cabinets, computer hard drives, and commercial data disks. As a result of this "search and fetch" activity, the need for a better-integrated data collection repository in the district was recognized.

The statistician gave each individual student an ID number, then created spreadsheets containing achievement data. This was a crucial piece of work, enabling the linkage of teachers' instruction in the LC framework with the students s/he instructed in a given year with student achievement on a variety of measures. Unfortunately, the time necessary to compile and merge datasets extended into much of the time budgeted for analysis.

Findings

Below are the findings. They are primarily quantitative and organized by research question.

Question 1: What is the relationship between teacher experience with LC and depth of implementation of the LC framework?

Information on the timing of LC training was obtained for 75 teachers who taught in classrooms from kindergarten to third grade, as well as special education teachers. This information was used to determine years of LC experience, ranging from 0 to 3 years. On average, teachers in the district had 2 years of LC experience by the end of the 1999-2000 school year.

Implementation of LC Training. Within this population of teachers, 62 completed the self-report survey on depth of implementation — an 83 percent response rate. LC coaches were able to provide observational ratings on depth of implementation for 61 of the teachers, although not on all facets of the LC framework. There were cases of missing data for items where coaches did not observe teaching practices.

Scores were computed for six domains of the LC framework (with items for each domain averaged together): writing workshop, guided reading, word work, shared reading, interactive writing, and read-alouds. Tests of internal consistency were conducted for both teacher report and LC coaches' ratings to assure that these scores were reliable. For LC coaches' ratings, Cronbach alpha scores showed strong internal consistency for all components, except for Word Work (see below). In a few cases, deleting questions about materials made a slight difference, but not enough to warrant deletion altogether. Internal consistency was not as good for teacher reported implementation, suggesting that the coaches' ratings may be more reliable.

To increase triangulation, 10 teachers were observed and rated a second time by LC coaches. Of the 10 cases, raters were in perfect agreement on only two teachers. In addition, the second set of raters, being in the classroom for limited amounts of time, had a higher ratio of missing data (54%) than observed ratings. Of the items rated by both LC coaches, there was perfect agreement on 76 out of 137 items (55%). Strategically, getting these second set of ratings was also a challenge. Although the 10 teachers were randomly chosen, they felt singled-out and unfairly scrutinized. Great care was taken in explaining the need to test for reliability and to persuade them that this piece of data collection was not a reflection on their teaching ability, but rather on the LC coaches' consistency in rating LC implementation.

As a group, teachers tended to rate themselves highly on all aspects of LC implementation, while LC coaches tended to be somewhat more critical in their observations. In almost every case, teacher self-report scores were higher than were LC coaches' evaluations. In addition, the range of teacher response was narrower (for most

questions between 4 and 6), while it was broader for the LC coaches (more scoring at the lower levels). (See Table 1.5). Even though LC coaches tended to give lower ratings in their observations, these ratings had strong positive correlations with teacher self-reports in three areas: Writing Workshop, Guided Reading, and Shared Reading, and a weak correlation in the domain of Interactive Writing. (See Table 1.6)

Based on the analyses, we posit that the LC coaches' ratings are a more reliable measure of depth of implementation based on: (1) Cronbach alpha results that were stronger for the coaches' ratings; (2) close or perfect agreement in many instances between coaches in our tests for interrater reliability; and (3) the increased variability among the coaches' ratings compared to teachers' self-reports. Although these findings are somewhat mixed, together they make a compelling argument for placing greater emphasis on the coaches' ratings.

Table 1.5
Average component scores for LC implementation

Component **(Cronbach alpha for LC Coaches Rating)**	**Teacher Self-Report** **Mean (s.d.)**	**LC Coaches Rating** **Mean (s.d.)**
Writing Workshop (.80)	4.8 (1.01)	4.4 (1.25)
Guided Reading (.86)	4.7 (1.24)	4.1 (1.57)
Word Work (.67)	4.8 (1.16)	3.3 (1.46)
Shared Reading (.85)	5.1 (.89)	4.7 (1.30)
Interactive Writing (.92)	4.9 (.93)	4.1 (1.43)
Read-Alouds (.85)	4.85 (.92)	4.85 (1.30)
Frequency of professional development	1-3 time per year on average	N/A
Value of LC Experience	(more than) somewhat valuable	N/A

Correlations were used to test the relationship between years of teacher experience using the LC framework with both teachers' self-report of implementation and LC coaches' observations of implementation (See Table 1.6). The results revealed that years of LC experience was not significantly correlated with depth of implementation for any of the

six elements as measured by teacher reports or LC coach ratings. (Although a positive correlation between teacher-reported depth of implementation of Read-Alouds and years of LC experience approached significance). The trends between teacher reports and LC coaches' rating were discussed at length, especially given that some of the strongest staff rated themselves lower than less skilled staff. It is a struggle to explain these confounding and surprising results. The belief is that it reflects a tendency for more experienced LC-trained teachers to embrace the reflective component of the program and be more self-critical. This phenomenon of being "unskilled and unaware" has been recently studied by Kruger and Dunning (1999), which showed that individuals need a certain threshold of knowledge in a particular domain in order to accurately assess their skills in that domain.[3] Furthermore, significant negative correlations were found between years of teaching in general and LC coaches' assessment of teachers' implementation of Shared Reading and Interactive Writing (and a negative trend overall). This may suggest that more veteran teachers have established teaching methods in these areas and may be resistant to new LC strategies.

Table 1.6
Correlations between teacher reports and LC coaches' ratings of depth of implementation

	Writing Workshop	Guided Reading	Word Work	Shared Reading	Interactive Writing	Read-Alouds
Teacher-Coach Correlation	.54***	.55***	-.16	.49***	.26~	.15
Teacher Report Correlated with Years of LC Experience	.04	.03	.15	.16	.02	.25~
LC Coach Rating Correlated with Years of LC Experience	-.01	.01	-.05	-.18	-.14	-.14
LC Coach Rating Correlated with Years of Teaching Experience	-.25~	-.21	-.19	-.30*	-.25*	.03

~p<.10 *p<.05 ***p<.0001

[3] Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology*, 77(6), 1121-1134.

Question 2: What is the relationship between a teacher's depth of implementation of LC and student achievement in reading and writing?

Achievement Outcomes for Students Taught by LC-Trained Teachers. Initially, reading and writing outcomes were reviewed by school and by grade for patterns of achievement in the district. For each year of data, students in classrooms with LC-trained teachers were tested to see whether they scored higher than students in classrooms with non-LC-trained teachers (See Table 1.7). Teachers were not randomly assigned to training, but initially invited to volunteer. In subsequent years, LC training became mandatory with teachers given the option of taking the training either within two years of the date of mandatory implementation or in their first or second year of hire. There may be some difference in teachers' implementation related to their enthusiasm for taking the training, but this was not explored.

Table 1.7
Average reading / writing assessment scores, 1998-2000

Variable	Teachers LC Trained			No Training		
	1998	1999	2000	1998	1999	2000
Kindergarten						
Percentage Score		0.68 (n=166)			0.63 (n=60)	
First Grade						
Instructional Text Level	2.02 (n=248)	1.75 (n=205)		2.03 (n=117)	1.41 (n=32)	
Text Book Level	1.8 (n=241)	9.68 (n=239)		1.67 (n=111)	8.00 (n=35)	
Second Grade						
Instructional Text Level	2.58 (n=146)	2.04 (n=140)		2.56 (n=143)	1.87 (n=61)	
Text Book Level	3.25 (n=141)	14.77 (n=201)	14.10 (n=246)	3.06 (139)	14.19 (n=88)	12.38 (n=32)
Third Grade						
Jerry Johns			4.62 (n=89)			4.27 (n=194)
Writing Assessment Score			5.08 (n=118)			5.74 (n=165)

Table 1.7 suggests differences in the average reading and writing assessment scores, with LC-trained teachers' students, for the most part, achieving higher average assessment scores than students in non-LC-trained teachers' classrooms. However, the positive

differences are not significant for all of the years. For the 1999 outcomes, t-test results suggest that kindergarten students in classrooms with LC trained teachers have slightly better scores, with a difference that approaches significance ($t=1.86$, $p<.06$). The t-test results suggest that first grade students in classrooms with LC trained teachers have better scores for Instructional Text Level ($t=2.56$, $p<.02$) and slightly better Text Book Level scores ($t=1.92$, $p<.06$). It should be noted that the large jump in scores for Text Book Level between 1998 and 1999 for first and second grade is explained by a difference in metric of the instrument, and not by teaching practices or cohort. LC and non-LC should be compared to each other by year, rather than comparing one year to another.

For the 2000 writing outcomes, there is a different finding. In this case, students in third grade classrooms with LC trained teachers got significantly lower writing assessment scores than students in non-LC classrooms ($t=2.85$, $p<.01$). However, surveys suggest that the framework may not have been as applicable for third grade classrooms so these differences may reflect school differences rather than teacher differences.

Student Achievement Outcomes and Depth of Implementation. A series of correlational analyses were conducted for first, second, and third grade reading and third grade writing outcomes for the 1999-2000 school year, concurrent with depth of implementation scores. Depth of implementation, as measured by both teacher self-reports and LC coaches' observations, was found to have a positive significant association with first grade running record scores only. Across the six separate elements of implementation measured by LC coaches, weak but significant positive correlations existed for writing workshop [$r(277)=.25$, $p<.0001$], guided reading [$r(277)=.31$, $p<.0001$], word work [$r(277)=.21$, $p<.001$] and shared reading [$r(277)=.21$, $p<.001$]. Correlations for interactive writing and read-alouds with first grade running records scores approached significance [$r(277)=.11$, $p<.07$] and [$r(260)=.11$, $p<.07$], respectively. The global score of depth of implementation across all domains was positively related to first grade running record scores [$r(277)=.24$, $p<.0001$]. Tests for second and third grade outcomes showed no significant positive effect of implementation, for either elements or global scores.

For comparison purposes, correlation results for both the coach ratings and the teacher reports are presented in Table 1.8 below. The coach ratings show stronger evidence of the relationship between Text Book Level scores and depth of implementation. But for reasons stated previously, more emphasis is placed on the coaches' ratings because they appear more reliable measures of implementation.

Table 1.8
Comparison of depth of implementation scores to text book level scores

	Writing Workshop	Guided Reading	Word Work	Shared Reading	Interactive Writing	Read-Alouds
First Grade Text Book Level Correlated with Coach Rating	.25***	.31***	.21***	.21***	.11~	.11~
First Grade Text Book Level Correlated with Teacher Report	.20***	.31***	-.07	.01	-.04	.20***

~p<.10 *p<.05 ***p<.0001

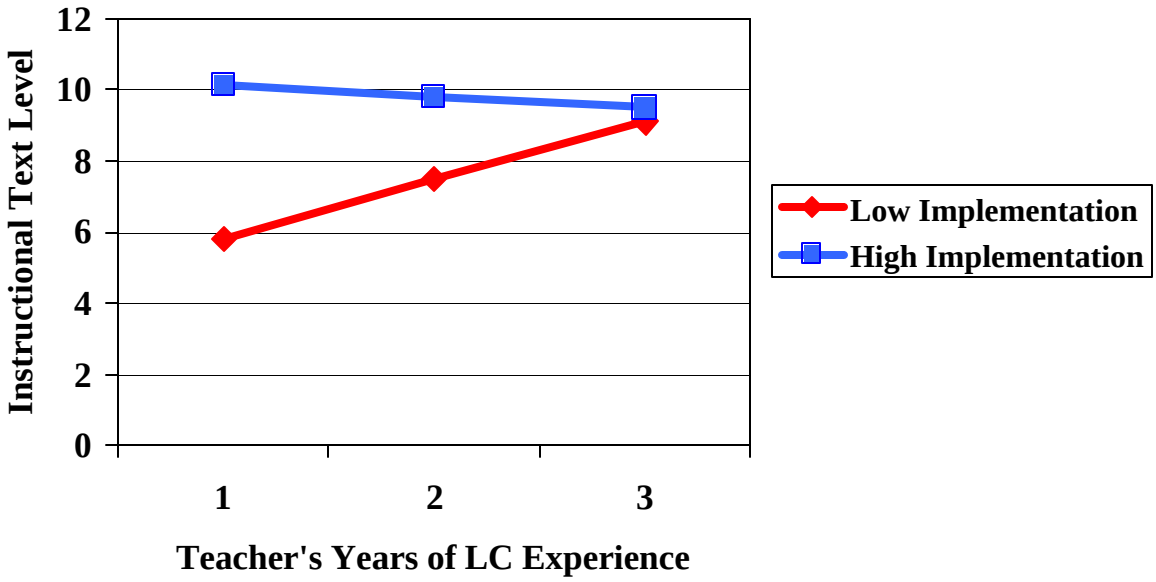
Question 3: Is the K-2 Literacy Collaborative an effective professional development model to improve student literacy?

Quantitative Findings-Regression. Given the positive results for the first grade cohort Instructional Text Level scores (1999-2000), a series of multiple regression models were fitted to estimate the predicted effect of teachers' years of LC experience and depth of LC implementation, controlling for students' socioeconomic status as measured by free or reduced price lunch status. Parameter estimates for years of teachers' LC experience and for LC coaches' observations of depth of implementation were significant and positively related to reading outcomes. In addition, there was a significant interaction between years of teachers' LC experience and the LC coaches' observations of the teachers' depth of implementation. Free or reduced price lunch status was significant and negatively related to student reading outcomes.

The predicted score for students' Running Record based on these estimates (See Figure 1) were plotted using high (3^{rd} quartile=5.60) and low (1^{st} quartile=4.38) values of depth of implementation by years of teachers' LC experience for students with free/reduced price lunch status. On average, estimates show that children receiving free or reduced price lunch read at a little less than two book levels below other children. As shown in Figure 1, predicted scores were higher for students of teachers with greater observed depth of implementation. However, the effect of years of LC experience was moderated by depth of implementation, such that scores for students of teachers with low depth of implementation but maximum years of LC experience were similar to students' scores of teachers with high depth of implementation. Children in classrooms with teachers who have low implementation scores but three years of LC experience have a predicted reading level that is over three book levels higher than children in classrooms led by teachers with one year of LC experience and low implementation scores. In contrast, children in classrooms with teachers with greater depth of implementation have higher predicted scores at each level of years of LC experience. For instance, it is predicted that a child receiving free or reduced price lunch who is in a classroom with a newly trained LC teacher with low depth of implementation would be reading at about an F level book.

A child also receiving free or reduced price lunch but in a classroom with a LC teacher with low implementation and three years of training would be reading at about an I level book. Children in classrooms with teachers with high levels of depth of implementation would be predicted to be reading at a I or J book depending on the years of LC experience.

Figure 1
Positive effect of Depth of Implementation and LC experience on First Grade Instructional Text Level scores (2000) plotted by effect of Level of Implementation moderated by Years of LC Experience for Students Receiving Free or Reduced Lunch (R-Square=.18).



Quantitative Findings-Cumulative Exposure to LC-Trained Teachers. An LC exposure variable was created by first coding whether or not students were taught by LC trained teachers (1=yes, 0=no) for each year of outcome data (LC98, LC99, LC00). Secondly, the variables were added together for the three years of data to test the cumulative effect of LC on third grade students' Jerry Johns reading, writing, and ITBS outcomes (see Table 1.8 and Table 1.9). Third grade students' exposure to LC ranged from 0 (never having been in class with an LC trained teacher) to 3 years (always enrolled in class with an LC trained teacher).

Table 1.8
Third grade 2000 average Jerry Johns reading score and writing score by LC exposure

Years of Exposure	Jerry Johns Score (n)	Writing Score (n)
0	4.01 (17)	5.12 (17)
1	4.33 (48)	5.62 (45)
2	4.60 (89)	5.97 (88)
3	4.70 (56)	5.40 (60)

Although students with more exposure did slightly better on average on both the reading and writing assessments, analysis of variance tests (ANOVA) suggested that the differences were not significant.

Table 1.9
Third grade 2000 average ITBS scores by LC exposure

Years of Exposure	Core Total: National % Rank (n)	Language Total: National % Rank (n)	Vocabulary: National % Rank (n)	Reading Total: National % Rank (n)
0	33.29 (17)	24.24 (17)	47.82 (17)	42.82 (17)
1	38.96 (45)	34.51 (47)	45.96 (48)	43.13 (47)
2	50.48 (81)	46.13* (84)	54.33 (84)	53.21 (84)
3	43.11 (55)	37.36 (56)	48.66 (55)	45.91 (55)

*ANOVA tests results show that Language scores for students with 2 years of LC exposure do significantly better than students with no LC exposure ($F=3.81$, $p<.02$). Additional correlation tests showed a weak positive relationship between LC exposure and Language that approached significance ($r(204)=.12$, $p<.10$).

Results of tests of students' LC exposure show a consistent pattern of higher scores for students with two years of exposure compared to three years of exposure. Although this is a puzzling pattern, there are several possible interpretations. Students with three years of exposure may really be more like students with two years of exposure. That is, that first year with LC trained teachers was also the first year of the program's implementation and training may not have been as smooth as it was in later years.

Table 1.10 below summarizes the quantitative findings related to research questions 1 through 3.

Table 1.10
Tests for differences in achievement associated with LC framework.

Year/Grade	Test	Positive Direction	Negative Direction
Enrolled in classroom with LC-trained teacher- t-tests			
1998/1st Grade	Instructional Text Level	Yes	
1998/1st Grade	Text Book Level		Yes
1998/2nd Grade	Instructional Text Level	Yes	
1998/2nd Grade	Text Book Level	Yes	
1999/Kindergarten	Percentage Score	Yes (p<.06)	
1999/1st Grade	Instructional Text Level	Yes (p<.02)	
1999/1st Grade	Text Book Level	Yes	
1999/2nd Grade	Instructional Text Level	Yes	
1999/2nd Grade	Text Book Level	Yes	
2000/2nd Grade	Text Book Level	Yes	
2000/3rd Grade	Jerry Johns	Yes	
2000/3rd Grade	Auburn Writing Assessment		Yes (p<.01)
Accumulation of LC Exposure (0 to 3 years) – ANOVAS			
2000/3rd Grade	Jerry Johns	Yes	
2000/3rd Grade	Auburn Writing Assessment	Yes/Mixed	
2000/3rd Grade	ITBS Core Total	Yes/Mixed	
2000/3rd Grade	ITBS Language Total	Yes/Mixed (p<.02)	
2000/3rd Grade	ITBS Vocabulary Total	Neutral*	
2000/3rd Grade	ITBS Reading Total	Yes/Mixed	

* Scores fluctuate up and down by years of LC Exposure, so the results are not clearly interpretable, as opposed to mixed results which show both positive results by one value of years of LC Exposure and no difference for other values of LC Exposure.

Qualitative Findings Related to Research Question 3. Other data to answer this question came from the last section of the survey, which invited teachers to respond to the question, "To what do you attribute your students' success in reading and writing?" Forty teachers responded to this question. Teachers identified their commitment to teaching literacy skills (13 teachers), their enthusiasm (8), parental support (5) and their own years of experience (2). The Literacy Collaborative initiative featured even more prominently in teachers' responses. In 21 responses, teachers named parts of the LC framework or LC in general as their reasons for achieving success with kids. One teacher wrote, "Letter work, interactive writing (whole group and individual), shared reading, read-alouds, individual guided reading." Another wrote, "Building an environment that fosters the love for reading is the most important. I also use the many components of LC to support the skills involved in reading." A third merely exclaimed, "LC training! LC

framework!" Three teachers said the use of LC as a common language is a benefit in itself. Additionally, 9 teachers named the support they received in their LC training.

Teachers also commented on the survey itself (in the margins and at the bottom of pages), in addition to a space provided for comments at the very end of the survey. These comments addressed a wide range of topics. Several addressed the study and survey design directly. One teacher wrote, "I am excited about this study and I look forward to hearing the results." Another described her reaction to the survey, saying, "I have problems with all the wording [on the survey] like powerful, exciting— It seems the choices on this paper are either you are a slow and boring teacher or an always exciting, fast paced super one."

Teachers wrote about LC and particular parts of the LC framework as well, identifying concerns, gaps, pieces of the framework that seemed most effective for particular grade levels and special needs students, and ways that a particular strategy was or was not working. Most of the framework-specific comments were directed to guided reading and interactive writing. They wrote about continuing professional development with LC (both for it and against it). One teacher praised the support she received: "I have received good support in using techniques and adapting for my level." Another voiced an opposing viewpoint, "We do not need more continuing contact classes."

Recognizing that anecdotal experience can add dimensions to quantitative and qualitative data from the survey and student achievement data, the research team spent some time identifying stories about LC's impact and summarized them into the following points:

- A distinct change can be discerned in many teachers who were at first polite about LC, and who are now deep into its practice.
- There is talk about LC everywhere in the elementary schools, signifying that it has become a common language.
- Conferences within the state and on the national level validate and confirm the LC framework, which lends credibility to our work in Auburn.
- Kids are excited about reading, and there are fewer conversations among teachers about "reluctant readers."
- The lowest-achieving students are performing much better than they were 10 years ago.
- An unintended consequence of LC training and greater attention to student achievement has been an increase in special education referrals, perhaps prompted by teachers feeling more accountable for student success.
- The district, individual schools, and individual teachers are looking at student achievement data more regularly. In two schools, student reading levels are

displayed (anonymously) on a wall in a faculty room or conference room in the school, as a visual display that can trace gains over the course of a school year. Other schools are collecting monthly or quarterly information on reading levels, measured against grade level benchmarks and reviewed with teachers regularly. Slow progress children, as evidenced by assessment and reading growth, are becoming the center of discussions focused around developing additional support.

- The picture of reading development is clearer to teachers and administrators.

- A new breed of reader is visible in the schools: they may be "slow as molasses," but they have strategies for reading.

- In classrooms, reading instruction is more individualized than it used to be. Teachers seem more skilled at gauging a student's zone of proximal development.

- Teachers understand the difference between teaching words in context and pre-teaching words.

- Teachers are more explicit about literacy strategies, such as rereading to make sense of the text.

- Special education teachers are also taking the LC course and using those strategies in combination with other approaches (such as Wilson).

- There is more willingness to use an array of strategies to help students gain skills, and less attribution of student achievement to background factors such as socioeconomic status.

- Thanks to LC, there is more focus on instructional practice. In addition, this focus has emerged into a centralized collaborative process. Prior strategies in the district tended to be less cohesive and included an array of programs or structures developed independently of each other.

Conclusions

The study reinforced the Auburn School Department's commitment to the Literacy Collaborative. As the district's Title I coordinator, said, "We made a commitment for the future. We still believe in LC. We're pleased with the framework. The teachers who are trained will continue to grow in their skills. New people will always need to be trained."

When the series of tests conducted with the data are reviewed, a positive trend for students taught by LC-trained teachers is visible (Table 1.10). While there are few statistically significant findings, the pattern of results suggests a small but positive effect of the LC framework. In practical terms, we might expect that a student exposed to LC implementation would gain a one text level improvement in a school year beyond that of a student not exposed to LC implementation. These differences might be more

pronounced with a larger sample and more accurate measurement instruments. Although it is possible that the differences have been found by chance, due to the series of tests run, there is less likelihood of finding scores for students in LC classrooms consistently (if not significantly) better than for students in non-LC classrooms. Qualitative data also supports the continued emphasis on LC training.

The research team was enthusiastic about the study at every step of the way. The team created a PowerPoint presentation and shared information to the administrative staff, school committee, and elementary teaching staff in December 2000 at a faculty meeting at each school.

Findings from the study can be summarized as:

- Years of experience with the professional development framework was not significantly correlated with self-reports of implementation (nor with LC coaches', perhaps due to the phenomenon that "the more you know, the more you realize you don't know." But neither were LC coaches' more conservative ratings of teacher implementation correlated with years of experience. We originally expected that teachers with more experience using the LC framework would also be inclined to implement the program in greater depth. This was not the case, which suggests that further study is warranted to determine influences of motivation for teachers who engaged in more comprehensive implementation of the framework.

- For the most part, students in classrooms with trained teachers had higher reader and writing scores than students with non-trained teachers. However, the differences were not pronounced, and they did not hold for all years in all skill areas.

- Depth of teachers' implementation of the professional development framework had positive significant correlation with grade 1 running record (reading) scores. Weak positive correlations existed for other components of the framework: writing workshop, guided reading, word work, and shared reading.

- A multiple regression model for grade 1 running record data, which included both depth of implementation and years of LC experience, indicated an interaction between these two measures. More years of experience can boost the influence of lower levels of LC implementation. However, higher levels of depth of implementation predict better results, even for teachers with only one year of LC experience.

- Students receiving free or reduced lunch have a predicted reading level that is about two book levels less than predicted levels for similar students not on assistance, underscoring the negative effect of poverty on children's achievement. However, being in classrooms with LC trained teachers with

higher levels of depth of implementation can substantially boost predicted reading outcomes.

- Even with some positive trends regarding depth of implementation, the measures themselves are problematic. Positively skewed responses on many of the items suggest the possibility of either the difficulty in critical self-assessment or a lack of clarity or specificity in the questions themselves. Despite the best efforts of the team, comments from teachers about the items also suggest a need for continued item development – some questions were cited as confusing or leading to respondents. In addition, the LC coaches were not able to get complete assessments of all practices from their limited classroom observations. While increased opportunities for observation might help with measurement, it is cost and time prohibitive. A number of teachers were reluctant to be observed and felt singled out for review, possibly disrupting their regular teaching routines.

- Students with two years of exposure to trained teachers show a consistent pattern of higher scores than students with no exposure to trained teachers.

- While gains in student achievement have not yet met expectations, the district believes there is enough good news in the study to stay the course with professional development training of K-2 teachers.

The members of the Auburn School Department believe that everyone benefits from looking carefully at data, at trying to understand what is working well and what needs to be shored up. This study helped Auburn identify real needs for a better repository for student achievement data, more within-district expertise to make sense of the data, and more engaging ways to involve teachers in using data to improve instructional practice.

TECHNOLOGY LITERATE TEACHER
JASPER CITY SCHOOL SYSTEM
Susan Poling

In 1999, after two years of group workshops, the Jasper City School System found that the computer skill level of its teachers remained low and school computers were being used primarily for drill and practice or reward activities. The school system decided to develop and implement a two-year program to address this need. The first year, Phase I, concentrated on establishing a common level of computer literacy among teachers. This study looks at the effectiveness of the district's decision to replace group workshops for computer skill training with a year of one-on-one tutoring, as well as whether this approach to training was effective, efficient, and what changes took place in the classroom as a result. Online computer skill testing, surveys, interviews and technical support records to were used to measure changes in skills and behaviors. Findings indicated significant improvement in the faculty's computer skills as well as the frequency and quality with which teachers used technology. Based on the wide range of teachers' baseline skills, the method was found to be efficient in that all teachers received more training time than the state average of 8 hours, teachers who needed the most help got it, no class time was taken away from students, and no other professional development topics were eliminated or cut back to provide time for computer training. The findings also demonstrated that teacher involvement when students use instructional software was a key factor in its effect on student performance.

Introduction

In the past five years school systems in the United States have been entrusted with billions of dollars in grant funding for technology hardware, connectivity, software, and training. Such an investment begs the question "Are we getting our money's worth?" The complexity of school environments, the diverse ways in which computers can be used, and the rapid changes in the technology itself make this a difficult question to answer. Most agree, however, that if teachers do not know how to use computers, there is little hope of students reaping the benefits of this mammoth investment.

Between 1997 and 1999, the Jasper City School System (JCSS) networked every school and equipped every classroom with one or more computers connected to the Internet. This rapid deployment of equipment meant a whole new area in which professional development was needed. Administrators also realized that the time-consuming nature of technology training meant that it would be competing for professional development time with other established core curriculum and special needs topics.

Jasper City School System has three elementary schools, one middle school, one high school, and a center for severely challenged (i.e. non-verbal) students. It serves 2,678 students from diverse socio-economic backgrounds. Twenty-three percent are non-white and 28 percent qualify for free or reduced price lunch. Class sizes range from 28 in some high school courses to 18 in kindergarten classes. Jasper is a semi-rural community with a population of 14,111. After the loss of hundreds of coal mining and textile jobs in the last 15 years many of the residents commute to Birmingham, Alabama, which is just 45 minutes away.

The district had begun its technology training by offering inservice workshops in the established format used for other professional development topics. It quickly discovered, however, that training teachers how to use computers was a long-term and complex challenge. The purpose of JCSS's evaluation project was to determine if recent training strategies used by the system were effective. JCSS hoped to gather information that would inform its own programming and help other school districts design effective training programs and demonstrate the importance of data-driven decision making.

Research Questions

This study addresses the following research questions:

1) Has individualized computer skill instruction enabled teachers to operate computers effectively? Has individualized instruction proved to be a more efficient method of acquiring these skills than group training? Has basic skills instruction reduced "downtime"?
2) Does increased professional development activity lead to increased use of digital content for instruction? How much of the digital content is used as a primary method of delivering instruction? A supplemental method? An enhancement to instruction?
3) Have teacher attitudes regarding technology as a curriculum strategy changed?
4) Does the use of digital content for instruction affect student performance, specifically the SAT9 reading scores of low-performing students?
5) Does increased professional development activity lead to increased use of technology for self-directed professional development? Are teachers communicating with parents and students using modern communication techniques?

Description of Technology Literate Teacher Program

The Jasper City School System's Technology Literate Teacher (TLT) professional development program was created in an effort to bring technology into the district's overall standards-based reform efforts. Following a comprehensive program review in the spring of 1999, the system found that although teachers and students used computers on a daily basis, much of this use was spent with "canned" drill and practice software or with low-level educational game software. Upon reviewing examples of technology-curriculum integration practices that provoked higher-order thinking, administrators acknowledged that the system's teachers would need better computer skills if they were to use technology seriously for academic gains.

In response to the review, the system formed a committee consisting of the superintendent, the technology coordinator, teachers, principals, and curriculum administrators to devise a new training program. This committee investigated training plans used by other school systems in order to find successful models. They discovered that most other systems were also using group workshops with some mentoring for added support. Committee members also reviewed available literature on the subject beginning with the standards developed by the International Society for Technology in Education (ISTE).

ISTE had developed two different sets of National Educational Technology Standards (NETS), one for students and one for teachers (NETS•T). The later included a profile of the "Technology Literate Teacher." Such a teacher knows how to use a wide variety of multimedia software and peripheral devices. Further, this teacher measures the success of technology integration by student success in core academic content, rather than the quality of the documents and presentations themselves. This profile fit the long-term goals of the system's new professional development program and the committee subsequently named its new professional development program the Technology Literate Teacher program, or TLT.

Although some literature had noted that teachers were more apt to use software as a result of training focused on integrating technology into the curriculum rather than training on basic technology skills (Milken, 1999), Jasper's experience with "integration" workshops had more often than not left teachers more inspired than empowered. When teachers attempted to reproduce what they had learned in such workshops, their lack of basic computer skills usually interrupted teaching and often resulted in lessons being abandoned in frustration. This ultimately served to erode their confidence in their ability to master computers. As the committee identified the weaknesses of past professional development efforts, it found that the generally low and wide-ranging computer literacy among the faculty was the primary reason the group workshops were ineffective at producing a high quality of technology-curriculum integration.

As a result, the committee decided to spend one year concentrating on computer skills in order to pave the way for significant progress in technology-curriculum integration. The committee, which had adopted the name "Jasper City Schools' NETS committee, developed a two-phase strategic plan to improve the system's use of technology in the curriculum. Phase I would last one academic year and would focus resources on creating a uniform level of computer literacy throughout the faculty. As with reading literacy, literacy with computers meant that teachers would reach a level of mastery at which they would possess the basic skills needed to progress independently. The training plan would teach these technical skills in a way that made their usefulness obvious to teachers and allow teachers to quickly apply the skills to instructional practices. As teachers acquired computer skills, integration workshops would be offered in order to help teachers apply their skills.

Once the faculty had mastered a common set of basic computer skills, Phase II of the plan would concentrate on guiding teachers into how to use technology resources and activities to solve curricular challenges. A group of teacher-mentors would be created during the summer between Phase I and Phase II. From this point forward teachers would be expected to learn basic skills independently or with a mentor. New teachers would be expected to have basic computer literacy, but would receive an orientation to the system's available technology by a mentor. This study examines the outcomes of Phase I.

The NETS committee identified three major obstacles to providing effective training. Finding the time for the training was the first. Inservice days were often filled with activities to meet other professional development needs and were scheduled too far apart to facilitate any real progress in computer literacy training. Pulling teachers out of class was not considered an option. The second obstacle was the faculty's wide range of skills. Group training had proven

ineffective due to the degree of variation. The third obstacle was teachers' attitudes. Some teachers still placed little value on technology.

The committee hoped to overcome these three barriers through the use of an appropriate training format. Several options were examined. Continuing a program of group workshops involving individuals with a wide range of computer skills was rejected because it consistently left some participants bored and others overwhelmed. Tiered group workshops, where individuals would be grouped according to their skill level, were also rejected because scheduling these presented too many logistical problems. Individual online training was rejected as too intimidating for the users who needed training the most as well as being difficult to monitor. After debating the merits of various possible formats, the committee concluded that one-on-one personal training had the greatest potential of providing successful computer skills training.

The resulting program included individualized computer skill training in the classroom during the regular school day with a professional trainer and curriculum-integration workshops after school and in the summer. Although TLT was implemented to solve some immediate problems, the committee designed it to be renewed annually as competency levels rose, new technologies emerged, and new goals were set.

ISTE's National Education Technology Standards were used as a guide in determining which skills would be taught. Jasper's committee trimmed and prioritized the list based on what software/hardware teachers had available and what could be accomplished in the time allotted for the training. Originally, a set of practice exercises was given to teachers at the end of each session. After several rounds of sessions, however, it became apparent that very few teachers found the time to use them. As an alterative, the technology coordinator began sending out carefully timed emails, file attachments, software, and websites needed by teachers in order to reinforce their training. This tactic was useful in challenging the teachers who thought they knew how to perform tasks but who really did not. In addition, it helped present real examples of how the skills being taught could be used.

TLT Program Components

Individual sessions paired a teacher and trainer for a 30- to 45-minute lesson in the teacher's classroom. The sessions were scheduled during the teacher's planning period or a pullout period such as physical education or music. This meant teachers lost no time with students and no substitutes were needed. The short session time allowed the pair to cover several skills while not overwhelming the teacher. The sessions were mandatory for all teachers except those who could demonstrate mastery of at least 85% of the NETS committee's list of essential skills. Ten teachers were able to do this at their first or second session, but only two opted not to continue the training. Principals excused several other teachers for various reasons such as coaching obligations. Originally each teacher was to have one session per month. The complexity of each school's calendar, however, only allowed for sessions about every six weeks.

After the individual basic skill sessions were underway, the plan added a set of group workshops developed around the teachers' newly acquired computer abilities. These sessions encouraged teachers to exchange ideas on technology integration and classroom management techniques.

They also allowed them to gauge their progress against that of their peers. Although the group workshops had been rejected as a format in which to teach computer skills, these sessions were carefully planned to use newly acquired skills in order to reinforce them and put them into a curricular context. By basing workshop activities on known skills, the teachers could have a successful workshop experience and leave with a viable project.

The topics, which were voted on by teachers, often involved the Internet and the use of peripheral devices such as digital cameras, digital projectors and scanners. The trainers, the technology coordinator, and fellow teachers taught these workshops. Workshops generally lasted one to four hours and were offered after-school, on inservice days, and in the summer. Attendance for group sessions was voluntary except in the case of inservice workshops.

In selecting trainers, the technology coordinator looked for individuals with computer training and education experience. Personality and the ability to make the connection between technology and core curriculum uses were key factors in the selections. Finding individuals who fit these criteria was a challenge, especially when the trainers were required to work full days, part of the time, with breaks for holidays and test weeks. Nonetheless, they were asked to make a commitment to the program so that the teachers could work with the same instructor over the course of the training. Unfortunately the first two trainers left after the first year to take other employment and were replaced with two new trainers, causing a delay in the training schedule.

Methods

This study collected information via surveys, professional development records, testing of teachers' computer skills, technical support request records, and teacher interviews. In order to assure teachers that this study was of the professional development program itself and not of their individual progress, each teacher was assigned a confidential identifier. These identifiers were used whenever teachers took surveys or skill tests.

Table 2.1 summarizes the study research questions and methods.

Data sources included the following:

Surveys

The teacher surveys were administered on computer. The faculty was emailed when the surveys needed to be taken and teachers took the surveys from their classroom computers where they would have access to professional development records. Survey responses were then downloaded into *Excel* for analysis. The teacher survey was administered in the spring of 1999, in the fall of 2000 and in the spring of 2001.

Parents and students were also surveyed to provide corroboration of the changes reported on teacher surveys. These parent and student surveys were administered on paper and responses were scanned. In the earliest grades K-3, the teacher read the questions to the students and tallied results by a show of hands. Older students took the survey during their homeroom period. Students took the parent surveys home at report card time. Each family was asked to return only

one survey back regardless of how many students it had in the system. The student and parent surveys were administered in the spring of 2000 and the spring of 2001.

Table 2.1
Research Questions, Data Collected, and Data Collection Methods

Research Question	**Data Collected**	**Data Collection Methods**
Has TLT enabled teachers to operate computers effectively?	Computer skills acquired	Interview, On-Line Assessments
Has one-on-one training proved to be efficient?	**Training Time, Computer Skills Acquired**	Records of Training, Baseline Scores, On-Line Assessments
Has training decreased support requests for "user-solvable" problems?	Technical support requests	Record keeping by technician
Did TLT lead to increased use of technology for classroom instruction? What type?	Hours per week digital content was used for instruction, enhancement, remediation, reward	Survey of teachers and students, Lesson Plans, Observations
Have teacher attitudes regarding technology as a curriculum strategy changed?	Teacher opinions on the importance of technology as a teaching strategy and motivator	Teacher Survey
Does the use of digital content for instruction affect student performance?	Reading test score data for low-achieving students	2000 and 2001 SAT9 total reading scores for 150 low-achieving students
Did training increase the use of technology for professional development?	Number of times per month teachers use technology for professional development	Teacher Survey
Did training lead to greater communication with students, parents, and community via technology?	Number of times teachers, parents, and students use email and web to communicate	Teacher Survey (Parent and student survey used as supporting documentation)

Professional development records

Teachers referred to their cumulative professional development records to provide information regarding their technology training. The cumulative professional development electronic form

teachers maintained listed any continuing education, workshops, conferences, or college courses they had attended during the past year.

Teacher computer skills testing

The original plan for assessing the TLT program was to have trainers record each teacher's mastery of skills on an individual progress checklist. The baseline score for 1999 represents the number of skills teachers were able to demonstrate to the trainer at their first session. When this study began, however, the decision was made to switch to *ActNow*, a performance-based computer skills training and testing software, in order to be sure data collected were objective.

To convert each teacher's mastery checklist to a number that could be compared with the *ActNow* scores, the trainers reviewed their original interview checklists and assigned each skill the same point values as given by *ActNow*. From that point on the teachers took *ActNow* tests. These were administered in the library or computer labs and were monitored by the media specialist. The test results were downloaded into *Excel* and the converted baseline scores were matched to the identifiers and added into the spreadsheet.

Scheduling of the baseline and the first progress check went well. The final round of skill testing, however, was delayed because the original trainers announced they could not return during the summer of 2000. It took several months to find suitable replacements, delaying the start of 2000-2001 training sessions by a month.

Technical support request records

As part of the normal maintenance process, teachers submitted technical support requests to the system technician on an electronic form including a detailed description of the problem or error message. The technician categorized the requests and stored the information in a database. During the course of the training plan, the technician kept track of all requests that an intermediate level user should have been able to solve. This information was presented graphically to determine if a relationship existed between the training and the "user-solvable" requests.

Teacher interviews

Teachers from each school were interviewed at the conclusion of the study to clarify findings and help to form conclusions about the effectiveness of the one-on-one training. Teachers were asked how large an impact the one-on-one had on their overall computer mastery, their ability to incorporate technology into instructional practices, and their motivation to use technology in core curriculum teaching.

Drawing the Sample

The system employed approximately 200 certified teachers at the time of this study. Because the study spanned two different school years, some teachers retired and left the study at the end of the 1999-2000 school year and new ones were added in the 2000-2001 school year. Table 2.2 reflects the tenure range of the initial group.

Table 2.2
Number and Percent of Teachers Participating in Baseline Year of TLT, by Years of Service, Fall 2000

Tenure	No. of Teachers	% of Teachers
1 – 10 years of service	45	33%
11-20 years of service	38	28%
20 and above	54	39%
All Teachers	137	100%

All teachers in the system have had computer access in labs since 1992 and in their classrooms since 1995. The Internet was connected to every classroom in 1997. In 2001, the system had 1,300 computers with a student to computer ratio of 3:1. Despite early access to computers, the drastic changes in computer operating systems during this same period minimized the benefits of early training initiatives. To help ensure the success of the training program, the system provided every teacher with at least one *Windows '95* computer in his/her classroom before the training program began in November of 1999.

Because this study did not begin until several months after the training had started, it was not considered practical to stop the training sessions for some teachers in order to create a control group. In addition, the system knew it would have no control over whether teachers learned computer skills outside the system's training plan. Therefore, all teachers who received training were asked to complete surveys and take tests.

The parent and student surveys were administered in the spring of 2000 and again in the spring of 2001. All parents and students were asked to complete surveys. Fifty-nine percent of students (1,531) completed the survey in the spring of 2000 and 52 percent (1,350) completed it in the spring of 2001. It is unclear what percentage of parents returned surveys because each survey could represent one or more students and the system did not have a clear number on how many households were comprised by its student population. Four hundred and ninety-two parents returned the survey in the spring of 2000 and 1,133 returned the survey in 2001. (The system has a student population of approximately 2,600). The results of these surveys were not matched to those of the teachers as this would have been a very complex task considering multiple children households and then students with multiple teachers. The results therefore, were used to show general trends and agreement between the teachers' responses and those of the students and parents.

Findings

The findings are organized by topics addressed in the research questions.

Has TLT enabled teachers to operate computers effectively?

The 1999 Milken Foundation Technology Counts Study (http://www.edweek.org/sreports/tc99/updates/states/al-t.htm) collected a wide range of information regarding technology in schools including hours of training, teacher computer literacy, and integration practices and paradigms. The study found that in 1999 nearly half (45%) of Alabama schools ranked the majority of their faculty at a beginner level of computer expertise. As seen in Table 2.3, teachers in the Jasper City School System fit this profile with 88% of teachers at a beginner (basic) or lower level of computer competency.

The system's TLT rubric, also shown in Table 2.3, called for a check of teacher skill levels midway through the training program. Because the training schedule had to accommodate holidays, standardized testing weeks, and diverse teacher schedules the midpoint fell in May 2000 and the endpoint in March 2001.

Table 2.3
Teacher Computer Skill Levels by Number of Sessions: (as in tables below)

Average No. Sessions	**Date of Assessment**	**Novice** (0-25%)		**Basic** (26-50%)		**Intermediate** (51-85%)		**Proficient** (86-100%)	
		Expected	Actual	Expected	Actual	Expected	Actual	Expected	Actual
Baseline	November 1999		77%		11%		11%		1%
3-4	May, 2000	0%	3%	10%	19%	80%	63%	10%	15%
4-8	February, 2001	0%	1%	1%	3%	70%	60%	29%	36%

The figures that follow demonstrate the teachers' progress throughout the training. Figure 2.1 shows the baseline scores of the original 192 teachers interviewed by the trainers. This graphic demonstrates the wide a range of computer competency held by the faculty. When these individuals came together in hands-on workshops, the disparity in their skills inevitably made the workshops unproductive and inefficient. This had stalled the progress in the overall technology program.

Figure 2.1
Percent of Known Computer Skills by Percent of Teachers, Base Level: October 1999

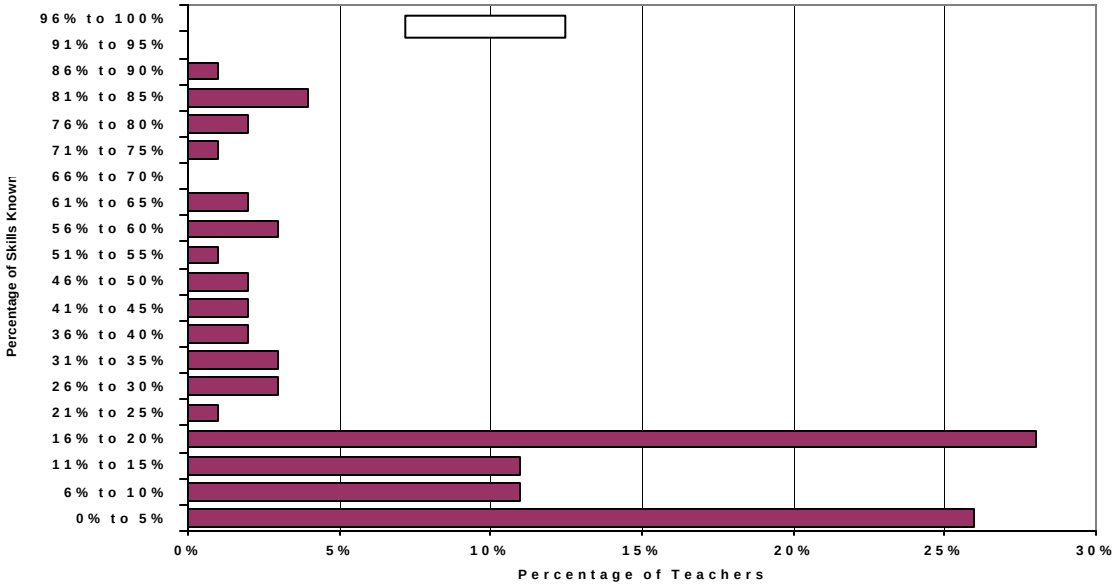


Figure 2.2 shows the faculty beginning to make progress as the training reached the midway point in May of 2000. The bell-shape curve that begins to form demonstrates that the faculty is assimilating the same information.

Figure 2.2
Percent of Known Computer Skills by Percent of Teachers, Assessment One: May 2000

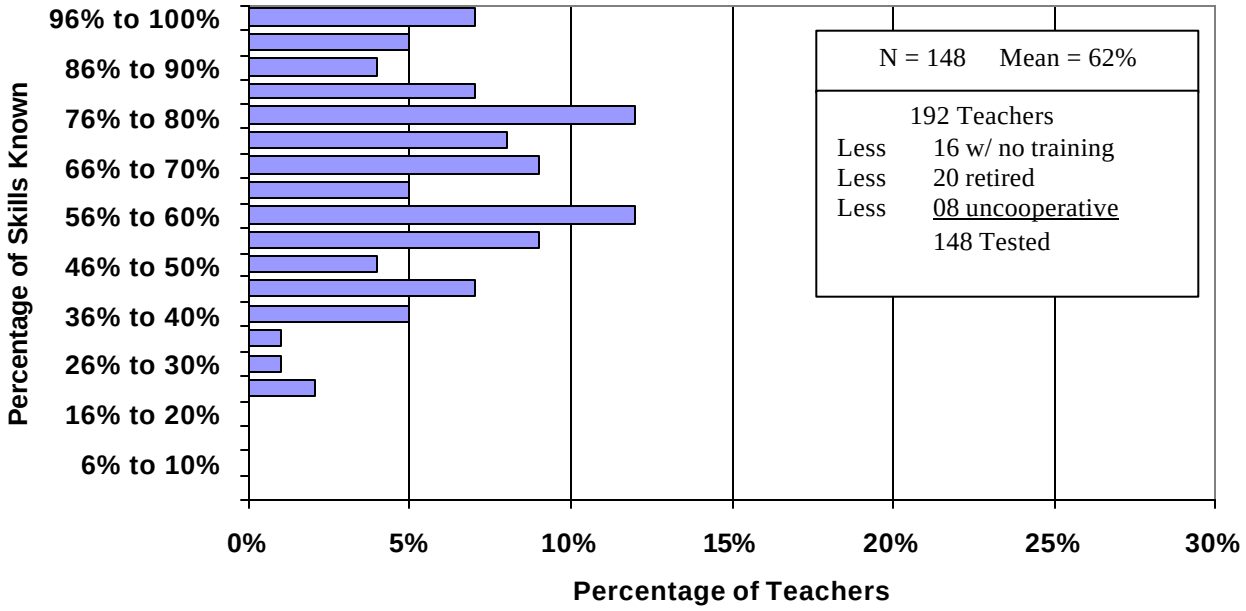
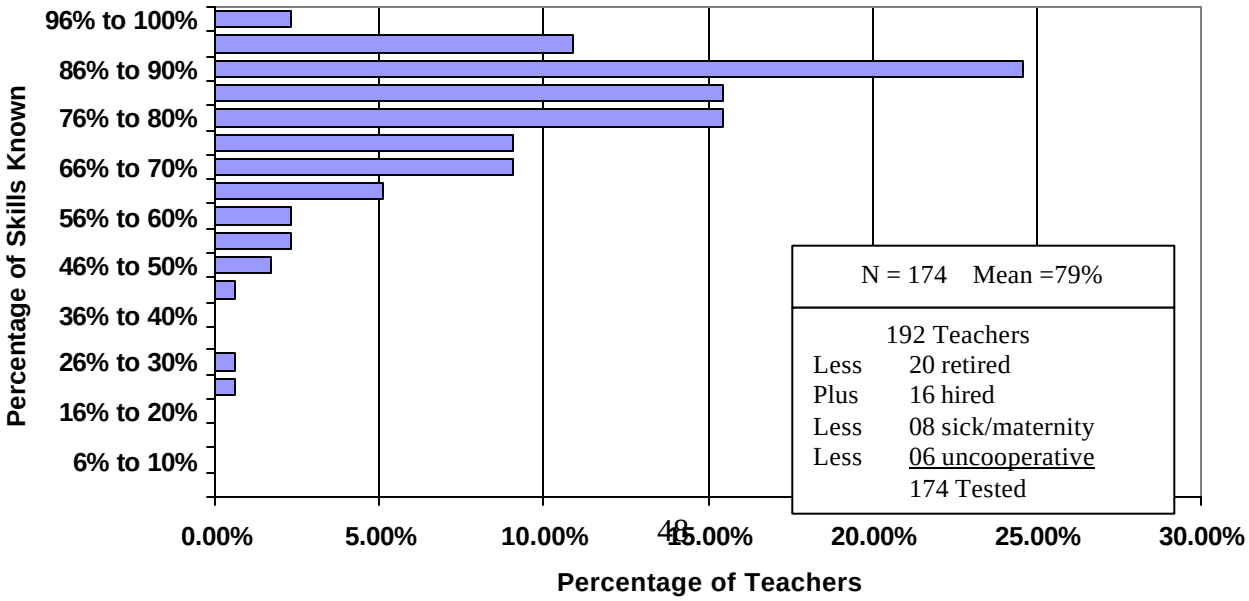


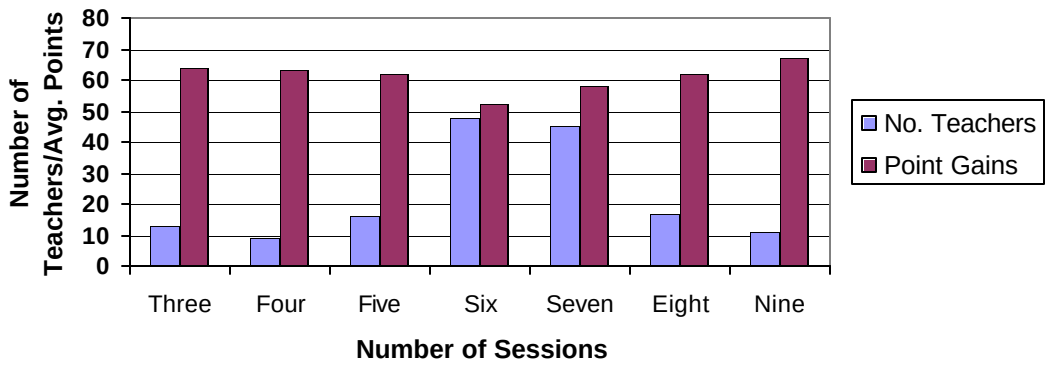
Figure 2.3 represents the final test scores recorded in March 2001. By this time, 96 percent of teachers tested had reached an intermediate or proficient competency level, knowing at least 51 percent of skills tested. As the distribution shows, the knowledge teachers had in common had grown, which led to much more productive group workshops.

Figure 2.3
Percent of Known Computer Skills by Percent of Teachers, Assessment Two: March 2001



A standard regression analysis revealed, revealed only a slight correlation between the absolute number of sessions and the amount of gains (improvements). This was not surprising given the fact that those who received the greatest number of training sessions were those that had the most difficulty learning the computer skills and those attending the fewest sessions needed the training the least. Figure 2.4 shows the teacher progress clustered by number of sessions.

Figure 2.4: Number of Teachers Participating in Multiple Sessions and Average Point Gain on Computer Skills Test, by Number of Sessions: 1999-2001



Self Assessment vs. Test Scores

Prior to conducting this study, JCSS had relied on self-assessments or on observations by the technology coordinator to evaluate the teachers' technology skills. In order to compare the reliability of self-assessments with the skill levels teachers were able to demonstrate on the tests, a self-assessment was added to the survey given at the end of the training program in March 2001. The terms used for the self-assessment were taken from a survey used by another school system and are shown in Table 2.4. The results of the self-assessment were graphed and then matched with the teachers' final *ActNow* computer skill scores. As Table 2.4 shows, teachers with a wide range of skill levels put themselves into all categories.

Despite the fact that the terms used for the self-assessment had more to do with actions than with expertise, teachers were surprised at these results when the information was presented to each faculty. They had expected more uniform correlations between test scores and self-assessment ratings. This information helped them to conclude that the test information was valuable and that it would be more helpful in planning future training than the results of self-assessments.

Table 2.4
Computer Skills Test Scores and Percent of Teachers Self Assessing Their Level of Technology-Use Ability at Various Levels, 2000

	Average Score	**Min./ Max. Score**	**Percent of Teachers**
Entry Level: just "getting started" learning how to use the available technologies.	78	65/89	11%
Progressive Level: continuously improving my skill levels and beginning to integrate technology into the curriculum.	75	60/90	64%
Proficient Level: have mastered the currently available technologies and routinely integrate technology into the curriculum	77	80/94	21%
Exemplary Level: experiment with new technologies and create innovative projects with students using technology	85	76/95	4%

Has one-on-one training proved to be efficient?

As stated earlier, creating control groups was not considered an option for this evaluation. The district could not keep teachers from learning on their own and, in addition, most faculty members attended one or more of the group "integration" workshops held during the year. However, as noted in Table 2.3, group training held in prior years had resulted in only 12% of the faculty scoring above a basic level of competency. Only after the addition of the one-on-one training did teachers in the system make any substantial gains in computer literacy.

The program's efficiency might also be measured in terms of the individual teacher's time. One of the main benefits of this training format was that the time each teacher spent in training was focused solely on skills they needed to learn. This eliminated time individuals spent in group workshops patiently waiting for others to catch up or being unable to keep up. In terms of facilitating overall professional development activities, the one-on-one sessions removed the need to schedule computer workshops on inservice days, making this time available for other topics.

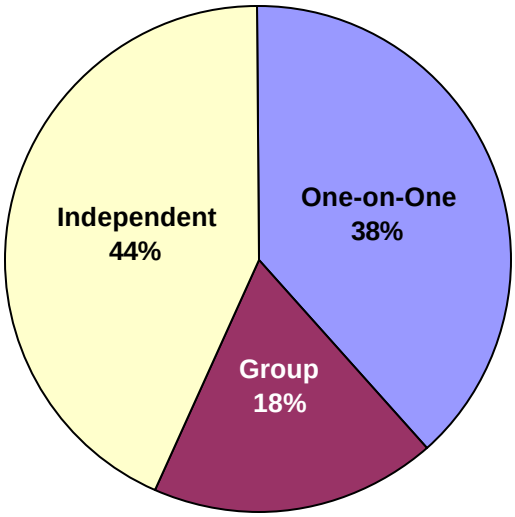
The one-on-one training format, however, presented other inefficiencies. Although teachers were continually encouraged to make the best use of their time with the trainer, this did not always happen. One first year trainer, who was also a psychologist, documented several ways in which teachers "blocked" their own training. For instance, teachers began sessions with questions about when new computers would be purchased or told the trainer what equipment

their children used at home. These discussions delayed focusing on the teacher's skill list. Trainers then had to steer the teacher's attention back to the specific computer skills without alienating them. While the trainers tried to keep the sessions focused on the skill lists, they were also charged with making each teacher feel comfortable, confident and motivated. Although this required dialogue that "wasted" critical training time, it had a positive effect on the overall outcome of the TLT program because teachers felt their individual needs had been supported.

Qualitative assessment of one-on-one instruction

At the end of the training, teachers were asked to identify what percentage of their new computer skills they learned in the one-on-one setting versus other settings. The results shown in Figure 2.5 indicate 44% of skills were learned independently.

Figure 2.5: Percent of Teachers Reporting Learning Basic Skills in Windows, Word and Netscape, by Format of Learning: March 2001



Subsequent teacher interviews revealed that teachers cited individual sessions as the critical element of the training program. The interviewed teachers felt that the independent learning was a direct outcome of the one-on-one sessions. In addition, they indicated that the short half hour sessions often left them needing more skills in order to complete projects, so they were also motivated to use Help files or other resources to find out how to complete and improve their work. In addition, teachers noted that the individualized sessions gave them the opportunity to focus on how to use their computer skills to target their teaching needs. Following are examples of how school staff viewed the training:

> *"When I sat down with the trainer in my room for the first time, she asked what I felt comfortable doing on the computer. I didn't even know what to tell her. I didn't have the vocabulary even though I could do certain things. As I became familiar with the basic skills and the vocabulary, I would just start finding what I wanted to do by using Help or*

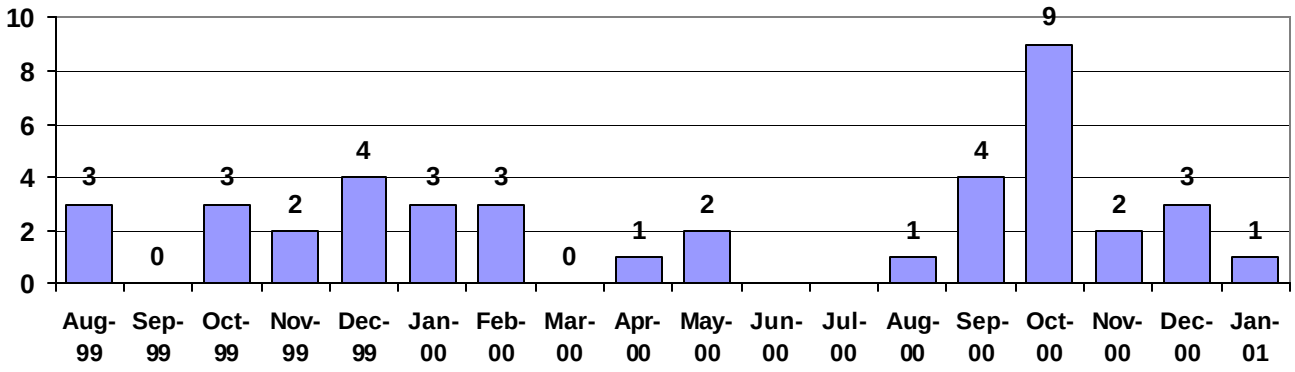
by trying new things instead of waiting for my next training session." Sixth grade math teacher

"This one-on-one training has gotten our teachers really motivated. I think we could have gone for years with training the old way in workshops and never gotten these results." High school principal

Has training decreased support requests for "user-solvable" problems?

It was hoped computer skill training would reduce the number of technical support requests for "user-solvable" problems. This was a goal not only to free the system technician for more complicated tasks, but also to eliminate the time wasted when teachers waited for technical support when none was really needed. According to the technician's records shown in Figure 2.6, the number of "user-solvable" problem requests did not decline over the course of the training.

Figure 2.6: Number of Technical Support Requests for User-Solvable Problems, by Month of Request: 1999-2001



Although the number of user-solvable problems did not decrease, the nature of the reported problems shifted to more sophisticated tasks. The actual requests made it clear that the TLT training simply "raised the bar." As users gained basic knowledge, they attempted more complex tasks that ultimately put them at the limits of their competency. For instance, complaints that the "network was not working" by teachers who were not logged onto the network were reduced to near zero; while requests on how to download "plug-ins" rose. This information has helped the technology planning committees to identify the need for school-based help desks that can provide "just in time" software support to teachers. Such a resource may help teachers implement technology faster and further reduce the amount of time the technician spends on non-hardware/network issues.

Did TLT lead to increased use of technology for classroom instruction?

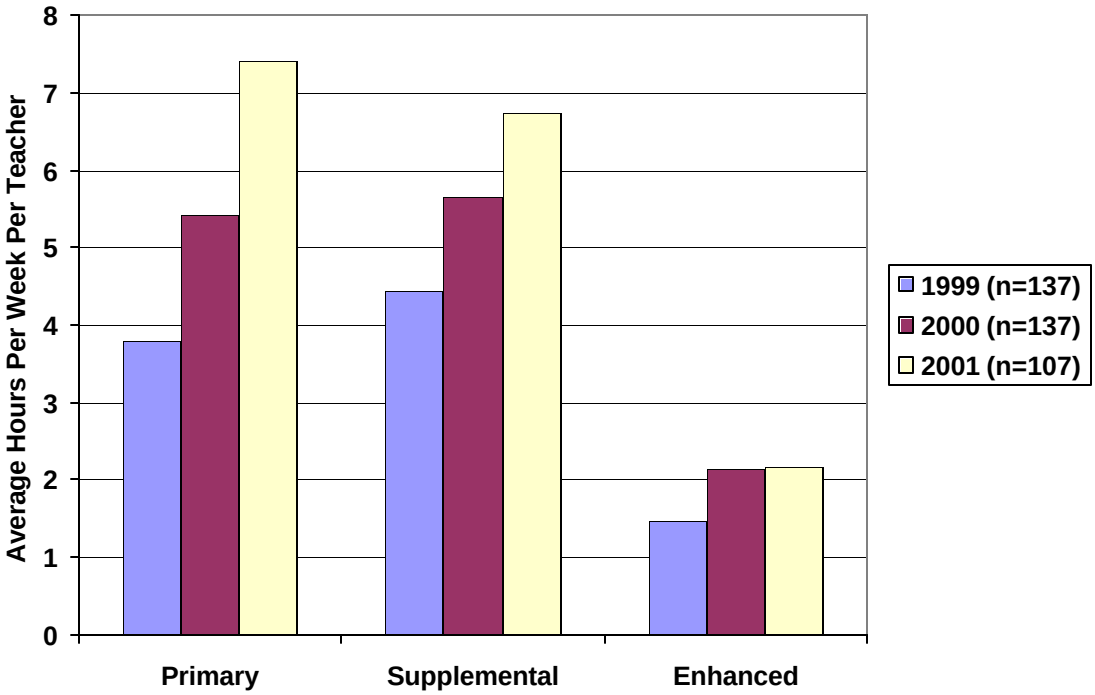
In order to measure changes in how teachers applied their new skills, teachers were asked how, when, and how much they were using technology. Teaching practices from the 1998-1999 school year, the academic year prior to TLT, were used as the baseline. The following definitions were used.

Primary Instruction: Technology is used by the teacher to present curriculum or by the student to engage in learning and research. The use of applications such as word processors and spreadsheets, when used as a substitute for pen and paper, were not considered a primary instructional use. The same applications, however, would be considered primary when they were used to teach or learn curriculum concepts. "Canned" educational software designed to present or remediate curriculum was not considered a primary instructional use.
Supplemental and Assessment: Technology was used to remediate, assess, or reinforce curriculum. This included using "canned" educational software. It also included the use of applications such as Microsoft Office when used to create documents as opposed to learned curriculum concepts. It could also include the use of *Accelerated Reader*, an online reading comprehension test, which is used pervasively at the elementary grades.
Enrichment: Technology used for rewards and non-critical enrichment of curriculum.

Data collected from the survey administered to all the classroom teachers show an increase in both primary and supplemental uses of technology over the course of the training. Uses of technology for enhancement appeared to stabilize as teachers gained new skills enabling them to make better uses of technology. This pattern of increase, seen in Figure 2.7, was the desired outcome of the training.

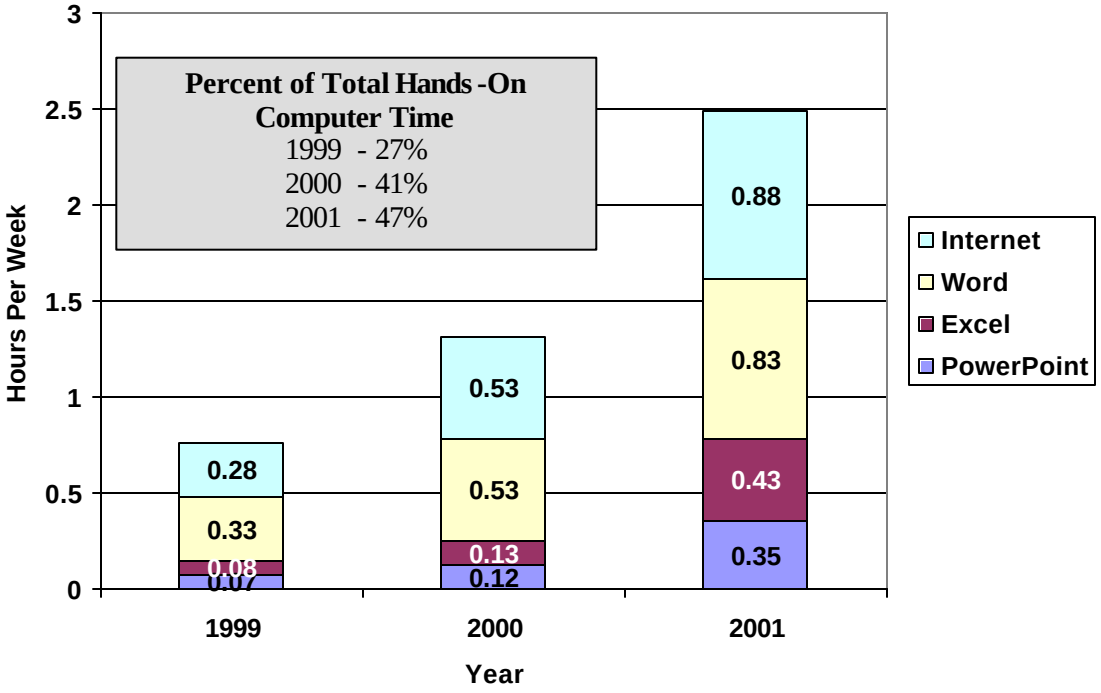
At the end of Phase I, each teacher was asked to create an "Electronic Lesson Plan." The written lesson plan had to focus on a core curriculum topic, identify the state or national standard it addressed, actively involve students with technology, and provide for an assessment of the content. The lesson plans provided insight into whether or not teachers were able to use their new computer skills and how creative they were in applying these skills to teaching strategies. (Note: This assignment was made after the spring 2001 survey was given. Therefore the increases seen in Figure 2.7 do not include its use.)

Figure 2.7: Teachers' Report of Average Number of Hours Per Week Computers Used for Instruction, by Type of Use: 1999 to 2001



In order to clearly distinguish between using "canned" drill and practice software and higher-level uses of technology, teachers were asked to identify how often students used general applications and the Internet to complete assignments during the school day. Figure 2.8 shows a significant increase in the use of these applications. A greater percentage of hands-on time, 47%, was devoted to this type of work in 2001, which indicates a rise in technology-based authentic work and research done in school. (Teachers did not routinely assign computer-based work to be done at home because only approximately half of the system's students have home access.)

Figure 2.8: Teachers' Report of Hours per Week Students Use Specified Computer Applications to Complete Assignments, by Type of Application and Year: 1999 to 2001



Have teacher attitudes regarding technology as a curriculum strategy changed?

Impact on learning

As shown in Table 2.5 below, teachers' ratings about the impact technology has on student learning rose steadily as the training plan progressed. This increase is seen as a direct result of the skills training that prepared teachers to use technology more effectively for instruction.

Impact vs. interest

Teachers consistently rated technology's impact on student interest slightly higher than it's impact on student learning, as shown in Table 2.5. This indicates that the lure of hands-on computer time may continue to provide a valuable strategy for engaging students in learning content that might otherwise be unappealing. The teachers' assessment of both interest and impact both rose at the same rate (.91) between 1999 and 2001. Once teachers have enough training to use computer technology more effectively in core curriculum, their opinions of it impact on instruction may rise more steeply than their regard for it as an attention-getter.

Table 2.5
Teacher Reports of Technology's Impact on Student Learning and Student Interest, by Year: 1999, 2000 and 2001

	1999 (n=137)	**2000 (n=137)**	**2001 (n=107)**
Student Learning	2.69	3.17	3.61
Student Interest	2.85	3.49	3.76

Note: 0=No Impact – 5=Large Impact

Does the use of digital content for instruction affect student performance?

To explore changes in student outcomes, the district explored teachers' use of a specific instructional software in reading.

In the fall of 2000, the system identified 150 students in grades 3-12 as reading below grade level. At the fourth grade level, in three different elementary schools, 36 students used instructional reading software in order to supplement classroom instruction. The students used the program three times per week for 20 minutes at a setting. The 2000 SAT9 Total Reading scores for these children were compared to the 2001 reading scores and showed an average gain of 1.2 percentile points. Other measures, such as a test with the STAR software program and teacher assessments, showed much higher gains and most all were considered as reading on grade level by the end of the school year.

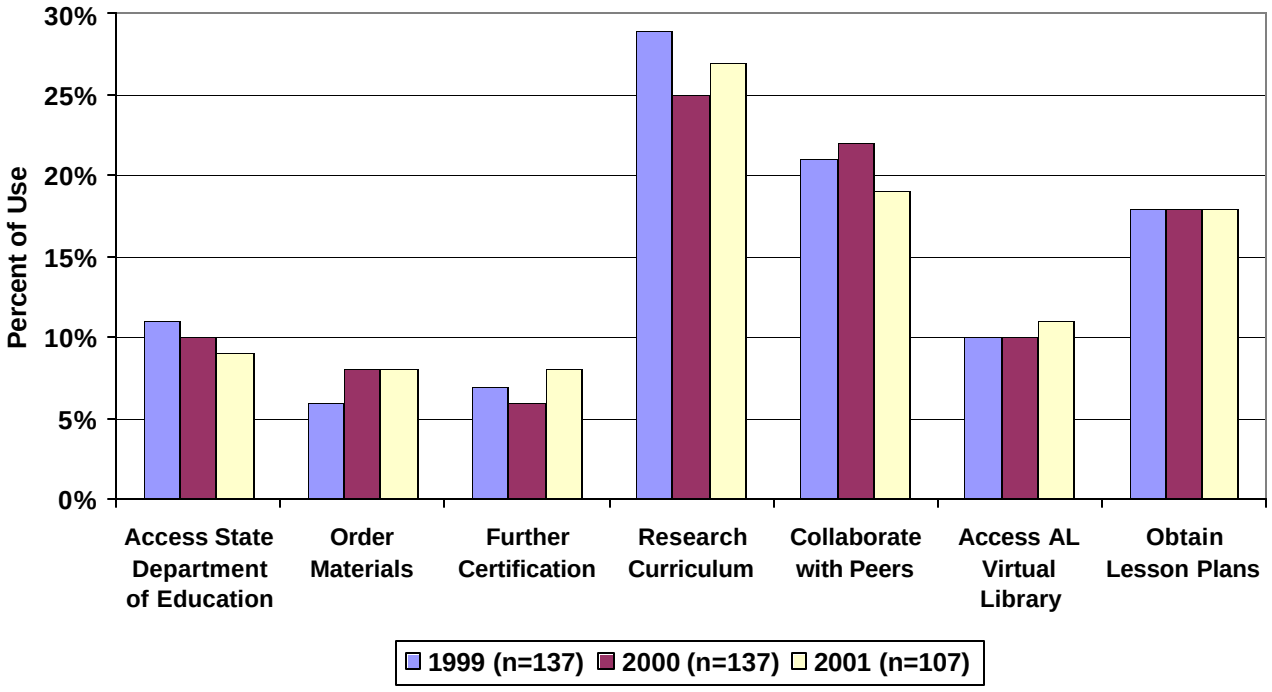
Each of the three elementary schools made its own decisions as to how students would access this software. All students used the software in a computer lab, but that is where the similarity ended. One school chose to have a teacher work with the students as they used the program. These students were given a great deal of direction and positive feedback. The 13 students at this school recorded an average gain of 11.8 percentile points in their SAT9 Total Reading scores. At the next school, the teacher accompanied the students to the lab, but left the management of the lessons to the lab aide. At this school, the students showed a 1.2 percentile point gain. At the last school, teachers sent the students to the lab and allowed a non-certified lab aide to supervise its use. At this school the 9 students who participated combined SAT9 Total Reading scores dropped an average of 11.6 percentile points. The important difference in the success of this program appears to be the teacher's involvement.

Did training increase the use of technology for professional development?

The average number of times teachers used online professional development resources/ communications rose from 6.68 uses per week in 1999 to 10.88 uses per week in 2001. Figure 2.10, however, shows teachers reported little change in the ways in which they used technology for professional growth. The areas that declined (accessing the state department of education, researching curriculum, and collaborating with peers) were a surprise. In the past year these three areas have become much more robust, user friendly, and pertinent. Two possible explanations for these declines may be the identity of the 30 missing respondents between 1999 and 2001 and

the teachers' changing interpretation of the terms used in the survey. With respect to the State Department of Education website, it may be that when teachers attempted to use it in the past, they found the interface too cumbersome and information available not worth the effort. In consequence, they may not have revisited the site.

Figure 2.10
Teachers' Report of Time Spent in Professional Development Activities,

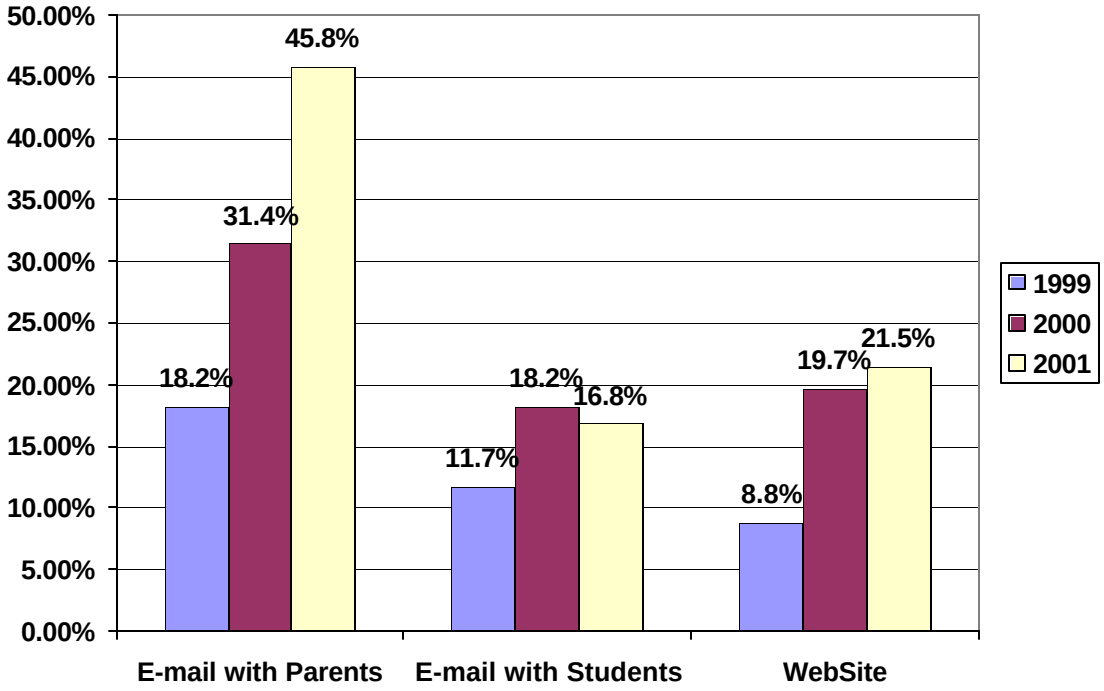


by Activity and Year: 1999, 2000, and 2001

<u>Did training lead to greater communication with students, parents, and community via modern technology?</u>

According to the teacher survey, e-mail between parents and teachers rose by 27.6 percentage points between 1999 and 2001. E-mailing students also increased between 1999 and 2001, but understandably occurred less often as students and teachers are in school together every day. Teachers reported e-mail as particularly helpful when students are out sick for any extended period of time. An increase in the number of teachers who created web sites to deliver information to parents, students, and the public also increased. This is an example of teachers obtaining outside training to use technology as e-mail skills had been included in the teachers' training, but creating web pages had not.

Figure 2.11
Percentage of Teachers Reporting Using Email and Web Sites to Communicate with Parents and Students: 1999, 2000, 2001



Parents and students also were asked if they used e-mail to communicate with teachers. Table 2.6 shows an increase in the percentage of students and parents who e-mailed teachers, which confirms that this form of communication was on the rise.

Table 2.6
Percentage of Students and Parents Reporting Use of E-Mail to Communicate with Teachers, 2000 and 2001

	2000	2001
Students	7.2%	10.9%
Parents	5.2%	6.8%

Conclusions

The findings of this study support the conclusion that the TLT training succeeded in increasing the teachers' ability to use computers and software. This improvement resulted in teachers gaining the confidence and ability to continue learning on their own, which is essential to keeping pace with the ever-changing computer industry. The amount of the gains leads to the conclusion that the individual training was a better method of instruction than group training

alone. The one-on-one format made it possible to teach each individual only what they had not mastered which resulted in the faculty gaining a common vocabulary and skill level so that they can work together productively.

Whether or not the expense of contracted trainers and the time required to organize and manage a one-on-one training schedule can be justified may depend on how broad the range of skills a school's faculty possesses and their attitudes. In situations where the faculty has a broad range of computer literacy rates and the preponderance of the faculty have low computer literacy skills, personal training may work better than online training. In addition, one-on-one attention may be beneficial in schools with a large number of reluctant users. Although Jasper's teachers had a generally positive attitude about computers when the program started, giving each teacher a positive learning experience with technology on a personal level noticeably raised interest and enthusiasm. Without some strategic commitment to raising all teachers' skill levels to an acceptable common level, a system may spend years holding workshops on beginning level skills.

Although the TLT program led to increased use of technology for instruction, this should not lead to the conclusion that successful training alone will result in such increases. Other factors may influence whether or not teachers use technology. For example, when JCSS teachers were asked to create the Electronic Lesson Plan at the end of Phase I, many teachers said they were reluctant to complete the project because of the time it required. Although having the necessary computer skills to complete a project may minimize the time needed to create a high quality technology-based lesson, it is still a time-consuming endeavor. Once the lessons had been created, however, many teachers admitted they had been intimidated by the project. Most reported being very proud of their lesson plans and were anxious to use the activities with their students. Developing self-confidence in teachers as it relates to computers was considered a necessary outcome of Phase I because it was considered a key influence on how often teachers would use technology in the future.

Another influence lies in the teachers' beliefs about technology. Some teachers continue to believe that instructional software will work effectively without direct teacher involvement. This relieves them from needing to know the how to use the program or from participating when the students use the software. The SAT9 results of the 4th grade students shown in this study provide some evidence that using technology in the absence of the teacher may be ineffective. The software developer had cautioned that the software would only produce good results when closely managed. The decision made at one school to allow this manager to be a computer lab aide rather than the teacher proved to be an unfortunate choice. Even at the school where the teacher worked with the lab aide (the lowest performing school), some progress was made. It was only at the school where the teacher participated fully in managing the students on the computer software that real progress occurred.

The findings showed that higher competency might not necessarily reduce the need for technical support. Teachers should, however, reach a level where they can distinguish between hardware/network problems and their own learning curve. The study's examination of technical support requests revealed that when support requests are documented, organized, and evaluated, they provide important information about teacher skill levels. The technician's database will

continue to be used to identify training needs by separating "user-solvable" requests from technical issues.

This study's comparison of the teachers' self-assessment with their skill test scores showed that teachers routinely under- and over-estimate their abilities. This should caution planners from basing their future professional development goals and strategies on opinion and brief observations. Technology training is such a time-consuming and long-term process that making the most of every professional development opportunity is key to making timely gains. In order to focus training where it is needed, highly detailed information is essential. As this study showed in how it conducted its baseline assessment, performance-based computer skill assessments can take place with or without expensive software.

Even when school systems can afford testing software they may be reluctant to ask teachers to take tests. Although Jasper's teachers were told that this study was being conducted to evaluate the TLT program and not them as individuals, many remained skeptical. The fact that the U.S. Department of Education was interested in the results made some feel more amenable and others more anxious. JCSS teachers responded very favorably, however, when shown the results of the testing. By showing the aggregated data in progress graphs as soon after the testing sessions as possible, teachers got a clear image of what it would look like to others and how the data would help future planning.

The results of this study helped to refine the system's plans for Phase II, using technology to strategically support the core curriculum. Although the individualized training sessions officially ended with Phase I, the system intends to continue supporting teachers individually in two ways. First, 20 teachers will be trained as trainers and mentors during the intervening summer in a program named Technology Academy. Second, high school students will staff a technology help desk at each school to assist teachers in implementing technology lesson plans and help them to continue learning new skills. In addition, one of the tasks of the Academy teachers will be to collaborate, document, and provide support aimed at reducing preparation time for technology-infused teaching.

Excerpts of this study were presented to faculty members in order for them to see their progress and make them aware of how the school environment is changing. A more complete presentation will be made to fellow technology coordinators at a symposium in the fall of 2001. The system's administrators, including the new superintendent, reviewed the results and seemed pleased with the progress made. As a result of this study, the technology coordinator intends to continue administering surveys and recording changes in classroom practices. The data acquired will be used to make future decisions regarding training, technical purchasing, and textbook adoption. It will also be used for public relations and to answer questions of the public and other stakeholders in the school system. As an extended consequence of this study, administrators plan to look closely at several other programs being adopted in a systematic manner.

CORE VALUES
OAKLAND UNIFIED SCHOOL DISTRICT
Heather Hughes, Laura Walker, Saul Rockman (Rockman et al)
Pamela Bovyer

This study explores middle school students' reactions to instruction from teachers who participated in an intensive professional development program. The training used technology using constructivist teaching methods to increase students' critical thinking and expertise in social studies and language arts. Evaluators investigated how instruction from trained teachers was received by students and how it affected student learning. To do so, evaluators interviewed students at six middle schools in classrooms of trained teachers, using open-ended questions. Students' responses were recorded and then coded. Results indicate that teacher practices changed in a variety of ways as a result of the professional development and that students consistently reported high motivation and a preference for constructivist lessons.

Introduction

Core Values is a middle-school curriculum and technology project implemented by the Oakland Unified School District (OUSD) in order to improve teaching and learning in social studies, English, and English language development classes. The Project seeks to support teachers as they integrate these traditionally distinct courses through intensive professional development, a curriculum-focused web site, and the distribution of computer technologies and other curriculum resources.

Core Values is designed to provide resources and training to support the numerous, systemic school-reform initiatives currently underway in Oakland in 14 middle and four K-8 schools. The curricular and pedagogical goals of the Core Values Project address the following four needs of middle school students, identified by OUSD: (1) *They don't think critically or historically; (2) They are not proficient readers, writers, and speakers; (3) They are not presented with opportunities to work within a constructivist approach to learning; and (4) They do not systematically use technology tools to support academic progress.*

Through policy initiatives, OUSD has laid groundwork for a "long-term, research-based, change process built around a constructivist pedagogy and rigorous content reform." Consistent with their research and the Department of Education report, "Caught in the Middle,"[1] OUSD has addressed the middle years with a series of policies aimed at improving student outcomes.

Oakland Unified School District has a median household income and the median house value are below average for the state of California. The district includes 54,763 students in 90 schools. The district employs over 2,500 teachers. Sixty percent of students in Oakland are eligible for free and reduced price lunch. The students in Oakland are 48 percent African-American, 26 percent Hispanic, 17 percent Asian, and 7 percent White.

[1] Fenwick, James J. (1987). "Caught in the Middle. Educational Reform for Young Adolescents in California Public Schools. Report of the Superintendent's Middle Grade Task Force. California Department of Education.

These policies include: school reconfiguration; block-scheduling and coring of language arts, social studies and English language development courses; and the adoption of performance standards in these disciplines consistent with California Academic Standards for Language Arts and State History/Social Science Frameworks.

While carefully constructed, these district policies and the timing of their adoption left teachers without curriculum, training or support. In particular, new teachers found themselves with no adopted social studies text, teachers trained to teach social studies were assigned to teach language arts, and few teachers knew how to teach in accordance with the new standards. Core Values was developed to mitigate these problems, and in partnership with numerous curriculum and technology organizations, such as the Bay Area Writing Project (BAWP), Teachers Curriculum Institute (TCI), Interactive University (IU) and Computer Strategies, the program aims to provide essential curriculum and technology expertise and professional development.

Since it began, Core Values has offered extensive professional development training in the use of Project technology and curricula to the more than 130 sixth and seventh grade teachers in OUSD. These workshops focus on writing and social studies instruction and technology application skills and curriculum integration. Training took place in eight-day summer institutes (for sixth grade teachers in 1998, seventh in 1999, and eighth in 2000) and weekly, curriculum-embedded technology professional development workshops during the school year.

During the initial phase of the program, each participating teacher received a TV/VCR unit, a cluster of four to five computers and a printer. As teachers attended related workshops, Core Values staff distributed peripheral technologies such as digital cameras, camcorders, and scanners. The Project also provided curricular materials and trained teachers to use them at Project workshops. Core Values built an extensive web site, hosting a variety of curricula, classroom activities, and links to primary sources online. Also, TCI curriculum materials were made available for those teachers who did not already have them.

Methods

This effort focused directly upon qualitative student impressions of the Core Values curriculum and pedagogy, which was supported through the professional development efforts of Core Values. The effort examined:

1) How instruction influenced by Core Values curriculum and pedagogy are received by students, and
2) How those practices, in turn, affect the way students learn?

Evaluators chose six middle schools with large populations of Core Values teachers. We asked for participation from all Core Values teachers at these six schools, a group that included teachers with varying experience using technology for instruction. We explained that we wanted to capture student impressions of Core Values curriculum and pedagogy. In soliciting participation, evaluators assured teachers that all information

would be reported only in aggregate; no names or school names would be used in reports. No teachers refused invitations, but a few never responded, resulting in a selection bias.

We instructed participating teachers to select a group of four students in one of their classes, two academically successful students and two less successful students. Students participated in a half-hour interview with an evaluator. Parental permission was required for students to participate. If students did not have permission forms or were absent on site visit days, then that teacher's group was smaller. Thus the selection of students also results in a bias, but involving teachers in their selection was less invasive to teachers. Each group of students evaluated curriculum in a one-on-one interview with an evaluator. Table 3.1 presents the research questions along with the data collected and data sources.

Evaluators obtained an outline of previously-implemented lessons from teachers and students, selected three-to-five concrete activities and, in individual interviews, asked students to comment on each one. Researchers aggregated student feedback, synthesized it, and presented the findings to each participating teacher who made time to meet with the evaluation team. Each teacher who received feedback responded briefly to the ideas generated by students, and their responses are included in this report.

During May and June 2000, we interviewed 94 students in 30 different classrooms from the six focus schools. In individual interviews at their schools, lasting approximately 25 minutes each, we asked students to comment on their Core Values coursework and activities.

Table 3.1
Research Questions, Data Collected, and Data Sources

Research Question	Data Collected	Data Sources
How Core Values curriculum and pedagogy is received by students?	Student responses to the following questions: (1) What they remember from their unit; (2) Which activities helped them learn and what they learned; (3) What they liked best/least about each activity; (4) How they felt about the teacher's and student's roles during the unit; (5) How they feel the rest of the class behaved and if others enjoyed the unit; (6) What they would change about the unit; and (7) What questions they still have?	Student Interviews
How those practices, in turn, affect the way students learn?	Student grades, test scores, and attendance	Student Records, including the district's Curriculum-Embedded Assessment (CEA)

Objective Student Outcomes: Standardized Test Scores, Grades and Attendance

When implementing any new curriculum or technology intervention, educators hope to portray the new strategy as one that will dramatically influence classroom practice and consequently students' academic achievement. Realistically, the treatment is often modest and the outcome measures are not clearly related to the intervention. In this effort, our analysis of objective student outcomes was inconclusive and warrants explanation of research limitations.

To assess any possible impact of Core Values on students, we originally proposed to look at student achievement by objective measures, including standardized test score, grades, and attendance. Unfortunately, there were problems with all of these measures. Grades are subjective; attendance was available once a day, not only during core classes and thus measures students' relationship to school more than to these particular classes; and available standardized test scores, the SAT-9 spelling, reading, language mechanics scores, focus on grammar rather than writing skills and analytical reasoning, which better describe the goals of the Core Values curriculum.

To supplement these problematic measures, we proposed to assess change in student writing samples from the Curriculum Embedded Assessment (CEA), a more relevant measure because it is directly connected to the Core Values and OUSD curricula, and measures writing skills, rather than mechanics. Unfortunately, only school-wide CEA data were available for analysis. Because Core Values involvement varied vastly by teacher, not school, school-by-school scores were insufficiently disaggregated to assess the impact of Core Values.

Thus, we used the institutional measures we had of grades and SAT-9 scores, not expecting to find much change. We applied our index of Core Values involvement to see if there were any trends among students who had *high use group* teachers in both sixth and seventh grade. Then we compared those trends to students of the other teachers who answered our survey and indicated less involvement in the program.

Unfortunately, only two schools had a sixth and seventh grade classrooms among the eleven teachers in the *high use group*. Between those teachers, there were fewer than twenty students in each of the three core areas of language arts/English, social studies, and English Language Development. Both schools are geographically in the flatlands, areas which face consistent achievement shortfalls next to the State and District norms.

Examining change from school years 1998-1999 to 1999-2000 in each of these measures was irregular between the groups. This small sample group's grades, test scores and attendance were often lower than other students. Due to the smallness of the sample group, we cannot determine statistical significance.

We made numerous other comparisons to explore and tease out any differences between the *high use* teachers and others in the Core Values Project, but we found no consistent or significant differences.

Findings

Each interview explored what students enjoyed in their classes, what they found educational, what they did not enjoy and why. Interestingly, in most cases, when students talked about what they liked and why, it often had to do with how much they learned. We asked questions about what students experienced in class and hypothetical questions, in which students described activities that were educational and exciting.

All of the students responded to open-ended questions, such as: What was your favorite project? Why did you like it? Why did you learn so much? What did you dislike? With the exception of questions about their computer use, and in some cases, group work, we did not ask students directed questions about particular aspects of their schooling. Therefore, it is quite possible that more students than we noted might appreciate, for example, hands-on projects, but did not express their opinions on the subject.

Students responded with great variation to the open-ended questions. To accurately portray their responses, we coded them into four different categories or types. Positive responses were coded as "like," "desire," "advise," and negative responses were coded as "oppose." When coding their answers, we clustered responses by topic, and identified them in tables by type of response. Definitions of the various response types are provided in Table 3.2 below.

Table 3.2
Definitions of codes used to classify students' responses

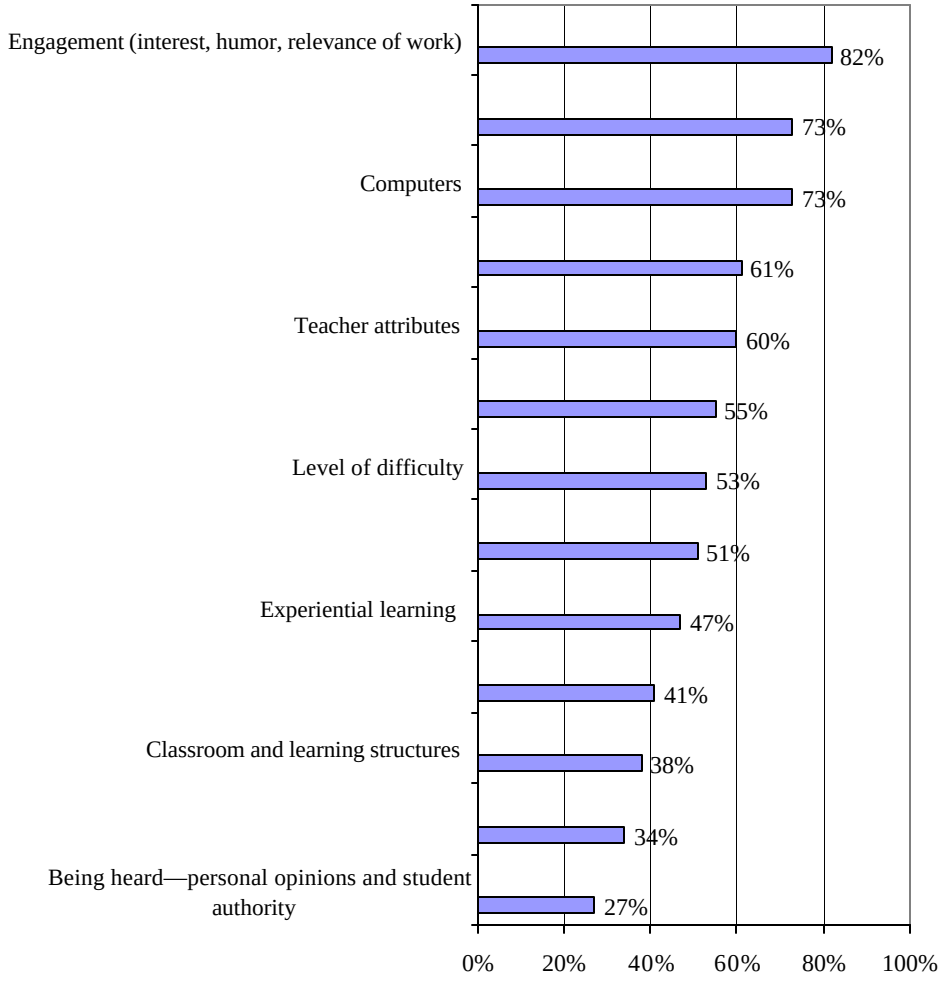
Type of Response	Example
Like: Students reported this practice in class and voiced appreciation for it.	Students appreciated getting to choose their own research Project topics in a recent class.
Desire: Students expressed a desire for a teaching practice or a dislike of its opposite, but recalled that their classes did not reflect this practice.	Students recalled an assignment and said that they wish they could have chosen their own topics, but topics had been assigned.
Advise: Students suggested ideal teacher practices and attributes. Students may or may not have been experiencing these kinds of classrooms when the data were collected.	In response to the question, "What would a great research project be?" a student describes an ideal lesson that involves student selection of research topics.
Oppose: These comments went against the grain. Students either expressed appreciation of the opposite or a dislike of practices popular among other students.	A student was frustrated when he had to choose his own topic for a paper.

In our analysis of the transcripts of the 94 student interviews, we created a list of the practices and/or activities that students mentioned as they discussed what makes their

classroom learning experiences positive. In Figure 3.1, these practices and activities are listed in order, beginning with the attributes mentioned by the most students.

For each topic of practices and/or activities that students mentioned, we have provided a table, indicating how many student responses there were within each of these four categories. Larger numbers in the "like" category indicate that classrooms are practicing this particular approach; larger numbers in the "desire" category indicate that classroom practice is not yet mirroring student desires. Larger numbers in the "oppose" category indicate ambivalence toward a particular practice for the group of students we interviewed. Numbers in the "advise" category show how often students suggested this practice in hypothetical situations. Note, however, that the same student may be counted in two different categories. For example, a student may have said he or she liked being able to choose his or her own topic for a paper ("like"), and that the student would allow for student choice if he or she were ever a teacher ("advise"). The text also includes indications of how many students raised the topics in their interviews whereas the tables (Table 3.3 through Table 3.15) indicate how many times various aspects of practices and/or activities were mentioned by the students. Because individual students often mentioned more than one aspect of practices and/or activities, the numbers in the tables frequently sum to greater than the number of students the text notes as having addressed a topic.

Figure 3.1
Student interviews: Practices and/or activities students mentioned as making classroom learning experiences positive by percent of students mentioning the practice and/or activity (N=94)



In the following sections, we discuss each of these practices and/or activities in greater detail. Because we wanted to preserve students' voices, each section is liberally peppered with representative student quotes.

Engagement

Students spoke about issues of engagement more often than any other aspect of classroom practice. Forty students talked about their interest, or lack of it, in a particularly curriculum area or assignment. Students praised assignments that they felt were interesting or relevant to them or curricula that they found engaging. Conversely, many students complained about having to learn something they found uninteresting. See Table 3.3 for details on the number of student responses regarding engagement. For instance,

("What was your least favorite part [of the unit on Africa]?") The salt Project…. because salt is boring—I didn't care much about salt…

Table 3.3
Student interviews: Number of student responses regarding aspects of engaging school-work that make classroom learning experiences positive

	Like	Desire	Advise	Oppose
Interesting work	19	14	12	0
Having **fun** while learning	21	0	20	0
Work that feels **relevant**	13	5	5	0
Other issues of engagement	11	9	7	0

When pressed to expand on what they meant by having fun while learning, students offered examples of motivating assignments. They included role play in drama, "making a game out of learning," or of a teacher being willing to joke around with students. To make lessons appealing, students advised that the process of learning and the content were both important. Some students were excited by content like "fun stuff on [Egypt],…studies about goddesses, and the Nile, …a project to draw a picture of Egypt, and a pyramid." Students talked about how novel experiences could be fun and help them learn.

Another important point was that students valued the importance of schoolwork that felt relevant to them.

(If you were a teacher what would you do?) I'd make things fun and relate them to how things are now so we can really know about them...

(discussing Internet research on China) I like writing on the Internet, it's more funny, and [has] more jokes; it's not all serious. I like stuff written by kids; I feel what they feel.

Students also discussed other aspects of engagement and meaning in their schoolwork. Eleven students mentioned doing something they're good at or in which they have been successful. Nine students talked about the importance of work that has a purpose. Others appreciated building up expectations and enthusiasm for a project before starting it, how respect in the classroom can be engaging, liking rewards and free time, and the appeal of being able to move freely around the classroom while working.

Projects: Building and Creating

Many students also commented on classroom projects in which they were able to build or create something. The vast majority of these students talked about projects in the last year

that they had valued and from which they learned a lot. Table 3.4 presents details on the number of student responses regarding projects (building and creating).

Table 3.4
Student interviews: Number of student responses regarding aspects of projects (building and creating) that make classroom learning experiences positive

	Like	Desire	Advise	Oppose
Building something	30	2	0	1
Creating something	36	1	5	2
Other issues of producing projects or products, or creating something	21	2	11	2

Building something. Projects in which students had to make or create something were extremely popular. Students talked about the pleasures of building something physical, whether it be a pyramid for a unit on Egypt, a castle for a Medieval Europe lesson, a Greek myth diorama, an African mask or a miniature island society. Many students explained why these projects were more educational than less hands-on assignments:

> *I liked building [a pyramid] ... you could put stuff in it and you could express yourself, and I had something to do with my hand; I like crafts.*
>
> *(Discussing a China report) I put mine on cardboard... I did the Purple Forbidden City museum... I showed the inside and outside and the walkways and I wrote the meaning that explains it... ("Did you learn anything?") Yes... because instead of the story itself, I could see what history is about.*

Creating something. Students praised opportunities they had experienced to create or design something, such as a drawing, story or game. Many students talked about why these opportunities taught them so much. "(discussing a creative writing project) I liked writing and imagining what might happen, and putting the story together with other people's stories."

Computers

Because computer use is central to the Core Values Project, this is one of the few areas where we asked students directed questions about their experiences. More than two-thirds (64) of the students we interviewed talked about how computers supported their learning. Nearly all students who commented on computer use were positive about it. Few discussed times when their use of computers was not educational, although some mentioned using computers to play games. Most student who commented on computer use talked about Internet research or made general comments about computer use and the

ways their teachers had used computers in their classrooms. Table 3.5 presents details on the number of student responses regarding computers.

Table 3.5
Student interviews: Number of student responses regarding aspects of computers that make classroom learning experiences positive

	Like	Desire	Advise	Oppose
Internet research	40	10	6	1
Computers are enjoyable generally	24	1	4	3
Gaining **skills**, typing	3	2	0	0
No problems with computers (amount of computers, glitches)	1	10	0	0
Other	14	4	5	5

Students were very positive about how computers impact their learning. They explained what they had learned. One student responded by highlighting the variety of information on the Internet, "You can learn about countries, people, birds, and animals and you can get pictures on it. It can also be fun; there are fun sites on it." Using Core Values curriculum, students reported using the computer to do research. They collected cultural and historical background on their projects, and then wrote up research reports. Students were impressed that the Internet provides more up-to-date information "faster" and they appreciated illustrations of ideas to supplement text.

Time: Too Little, Too Much

Fifty-seven students addressed issues of time during their interviews. In most cases, students were expressing a desire for change, rather than praising current circumstances. Said one student, ("What advice would you give to teachers?") Don't move too fast. Move at a medium pace. We had to learn about the Renaissance in two weeks….We had to learn about [it] in one week and then make a [Renaissance] newspaper in one week. It wasn't enough time. Table 3.6 presents details on the number of student responses regarding time.

Table 3.6
Student interviews: Number of student responses regarding aspects of time that make classroom learning experiences positive

	Like	Desire	Advise	Oppose
Adequate time to complete assignments	4	22	2	0
Particular activities are not too **time-consuming**	3	19	2	0
Variety of activities balanced well	4	17	8	0
Other issues of time	4	8	4	0

Teacher Attributes

More than half (56) of the students we interviewed brought up attributes that they liked or disliked about their teachers and which may have an impact on their learning. A number of students praised or wished for teachers who were helpful through their attitudes and actions, helping students understand and locate information or overcome difficulties. Table 3.7 presents details on the number of student responses regarding teacher attributes.

Table 3.7
Student interviews: Number of student responses regarding aspects of teacher attributes that make classroom learning experiences positive

	Like	Desire	Advise	Oppose
Teacher helps students understand, overcome difficulties	10	7	2	0
Teacher interacts with students (vs. desk sitting)	8	2	9	0
Teacher is nice/fun/happy, appears engaged, interested	6	3	4	1
Class is managed well, teacher has lesson plan	1	8	7	0
Teacher listens to students/respects their opinions	2	0	3	0
Teacher shows, rather than telling (discovery vs. lecture)	0	3	5	1
Other	8	7	6	0

Students also praised teachers who ran interactive classes rather than those who stayed behind their desks. Students suggested teachers lecture less and welcome student input more. One said, “They don’t call on me and make me wait a lot. I have a lot of

questions." They also suggested that teachers avoid using worksheets. Students who commented were divided equally among classrooms with teachers who exhibited the desired traits and those who did not.

Self-Direction

More than half of the students interviewed (52) touched on issues of self-direction. Most praised opportunities to choose their own research topics and agendas, while another significant group regretted the times when they had been told exactly what to do. One student asserted, "When you don't get to learn what you want, it's boring." Overall, students were excited about assignments in which they could make choices or explore the Internet or other research sources independently. One said, "(During a discussion of English notebooks: "What do you like least?") Having to write essays... thinking of ideas is hard, but if we got to choose own topics it would be more interesting and easier to write on." Table 3.8 presents details on the number of student responses regarding self-direction.

Table 3.8
Student interviews: Number of student responses regarding aspects of self-direction that make classroom learning experiences positive

	Like	**Desire**	**Advise**	**Oppose**
Choosing own areas to research/ learn about	12	8	4	1
Having **choices** in how to proceed/who to work with	13	2	6	1
Conducting **research independently**	7	4	0	0
Other issues of self-direction	14	6	8	0

Level of Difficulty

More than half of the students interviewed (50) discussed levels of difficulty with specific content, various activities, and teachers' instructions. Some commented that a particular assignment or activity was a good challenge for them—not too hard and not too easy. Some said they had requested a higher level of difficulty in their work while recalled lessons that were too difficult to be engaging. In general, students considered assignments that were too easy, boring, and those that were too difficult, overwhelming. Table 3.9 presents details on the number of student responses regarding level of difficulty.

Table 3.9
Student interviews: Number of student responses regarding aspects of level of difficulty that make classroom learning experiences positive

	Like	Desire	Advise	Oppose
Work is not too **easy**	3	6	3	0
Work is not too **difficult**	5	19	6	0
Students **understand** assignments, directions, etc	2	8	3	0
Other issues of difficulty	6	7	6	0

The following are some of the students' comments about difficulty with specific content activities and teachers' instructions.

> *("What did you think of writing down the questions today?") It was confusing… ("Why?") Because it was too many words I don't understand-- I really don't understand. ("What do you do when you don't understand?") Just sit there and don't do nothing.*

> *The funniest lesson I ever had was the culture and the atlas stuff I was doing yesterday. ("What made it fun?") It was fun because it was a challenge for me to look up all the countries and unscramble them. I like challenges.*

Learning Something New

Half of the participating students talked about the value of learning something new. Most of these students (44) recalled lessons and units in which they were able to learn new facts, explore curricular areas that they had not explored in previous years, or learn something unusual or different. Table 3.10 presents details on the number of student responses regarding learning something new.

Table 3.10
Student interviews: Number of student responses regarding aspects of learning something new that make classroom learning experiences positive

	Like	Desire	Advise	Oppose
Learning **something new** (e.g., new facts)	39	6	6	0
Work/ study areas are **not repeated** year to year	4	2	4	0
Learning something **unusual**	10	0	1	0

All of the students were extremely positive about these experiences. A few students also complained about lessons where nothing new was introduced, or content was repeated from the previous year. Several students spoke about the desirability of teaching something new when they were asked to imagine educational lessons and projects.

Experiential Learning

Almost half of the responding students spoke about the value they attribute to experiential learning, by which we mean non-traditional learning activities that allow students to experience some aspect of the topics they study. In this category we include movies, field trips, dramatic interpretations, guest speakers and authentic artifacts. Most of the students talked about the value they get from lessons when they can see footage of places they are studying. Those students who mentioned speakers or the opportunity to handle real artifacts recalled the experiences favorably. Field trips were extremely popular, either as real events or events that students wished would occur. Participating in plays around subject areas was also noted as a mostly positive activity, although a few students expressed dislike for plays. Table 3.11 presents details on the number of student responses regarding experiential learning.

Table 3.11
Student interviews: Number of student responses regarding aspects of experiential learning that make classroom learning experiences positive

	Like	**Desire**	**Advise**	**Oppose**
Movie viewing	13	1	11	3
Field trips	4	0	8	0
Doing **plays**	3	0	4	3
Having **speakers** visit	5	1	1	0
Seeing real **artifacts**	2	1	2	1
Other experiential learning activities	6	1	6	0

Interaction

More than forty percent of the students interviewed discussed the types of interaction and communication they preferred in their classes. Students praised the opportunities they had for discussions, presentations, and email and other computer interactions. Several students mentioned, however, that they did not like giving class presentations, typically because they did not like performing in front of a group. Table 3.12 presents details on the number of student responses regarding interaction.

Table 3.12
Student interviews: Number of student responses regarding aspects of interaction that make classroom learning experiences positive

	Like	Desire	Advise	Oppose
Discussion/gaining multiple points of view	8	4	1	0
Class **presentations**	10	1	3	6
Computer interaction (email, blackboard.com)	9	1	0	0
Other active participation (vs. passive listening)	9	1	10	0

The quotes below illustrate students' ideas about different sorts of classroom communication.

> *The most interesting [kind of lesson] is where you get to interact almost all the time; you get to express your opinion and talk about it, and do a little learning and talk about it some more, but you talk about it, not just all learning all day.*

> *("What are your worst ways to learn?") My worst ways are worksheets because you just sit there and she reads the answers and there's no studying… It's not interactive and she just tells us what to do...*

Classroom and Learning Structures

More than a third of the students mentioned classroom management systems and structures, such as managing of student behavior, organizing test preparation, providing study guides, and reading round-robin style. Most students discussed the ways in which the poor behavior of other students negatively affected their own learning experiences and expressed a desire for better managed classrooms. Several students who used interactive notebooks, a model for teaching students to take and use their notes, praised the ways the notebooks aided their test preparation. Some students objected to various learning strategies teachers used, especially having students read aloud. They commented that it was difficult to concentrate when others read too slowly or poorly. Table 3.13 presents details on the number of student responses regarding classroom and learning structures. One student said,

> *I learn the most by listening, and keeping my eyes on the teacher. Sometimes it's hard to see or hear, if someone is talking to me it's hard to listen to the teacher, then it's harder to learn from her.*

Table 3.13
Student interviews: Number of student responses regarding aspects of classroom and learning structures that make classroom learning experiences positive

	Like	Desire	Advise	Oppose
Other students' **behavior** that is conducive to learning	1	18	4	0
Students appreciate **learning structures** (test preparation, reading aloud, study partners)	10	1	1	6
Other issues of classroom and learning structure	1	0	0	3

Depth of Work or Information

About a third of the students we interviewed discussed issues of curricular depth—having access to the depth of information on a topic they desired or background information on topics that interested them. All those who commented regarded subject matter depth as positive. Most students were expressing a desire for more depth in their studies, but some students praised their current teachers for providing the depth of investigation they desired. Table 3.14 presents details on the number of student responses regarding depth of work or information. One student responded to the question, "If you were the teacher, what would you do differently?" as follows,

> *...like in history when she gives us a chapter to read, then next day she just collects questions and moves on, but I would discuss it a little more and give more background, if that's possible.*

Table 3.14
Student interviews: Number of student responses regarding aspects of depth of work or information that makes classroom learning experiences positive

	Like	Desire	Advise	Oppose
Students have access to **depth of information** or follow-up they desire	10	20	8	0
Other issues of depth	1	3	3	0

Being Heard – Personal Opinions and Student

More than a quarter of the students we interviewed brought up student authority and the value of voicing their own opinions. Most praised the opportunities they had in their classes to express themselves, either sharing their opinions on various subjects, creating projects that reflected their own thoughts and feelings, or sharing information about themselves (autobiographies). Others talked about assignments they valued in which they

were the expert, sharing their knowledge on a particular subject or teaching others. Table 3.15 presents details on the number of student responses regarding being heard.

Table 3.15
Student interviews: Number of student responses regarding aspects of being heard that make classroom learning experiences positive

	Like	**Desire**	**Advise**	**Oppose**
Students' personal opinions/ experiences count	14	1	3	0
Student is the expert/ teacher	7	0	1	0
Other issues of **student authority**	6	2	1	0

One student shared his ideas about having a voice and authority in class.

> *("What advice would you offer teachers?") Listen a lot to the kids. Then listen to their advice. Do something about it. For one day do it the way he asks. If it doesn't work, then do it the traditional way. Sometimes we get bored of listening, writing and doing work… We go to school five days a week, six and a half hours a day. It's good to have an hour of break each week*.

Conclusions

Taken together, these data suggest a few things about the implementation of Core Values, and the original research questions of, *how Core Values curriculum and pedagogy are received by students, and how those practices, in turn, affect the way students learn?*

An important premise is that Core Values was implemented variably. Among our sample groups, teachers who attended Core Values workshops and used the resources appreciated them and consequently had use for the ideas in their classes. The voluntary nature of the training was both its strength and weakness. Teachers who self-selected to participate extensively were very positive about it, but many teachers did not participate as extensively.

Students, exposed to Project curriculum, consistently assert that Core Values-types of lessons motivate them and help them learn. Students' comments support the efforts of the District to encourage constructivist, student-centered lessons that give students increasing responsibility for their learning. Whether students were appreciating project-based learning, advising for less lecturing or chastising teachers for worksheets, their feedback was clearly advocating for more efforts like Core Values.

When asked, our sample of students preferred constructivist lessons. Students want: fun lessons that are relevant; that allow them to use their hands; that involve computers; that

take the right amount of time to complete; that are challenging—but not too much so; that are experiential; where they can work with kind, helpful teachers; and where they can collaborate and talk to their friends. Most of these concepts parallel constructivist ideas and the theories behind the Core Values Project.

The value of the evaluation exercise exists in capturing student voices. Overall, students were excited to share their feedback about teachers' lessons. Their feedback seemed balanced and relatively consistent among students with the same teacher. Although a few students were obviously anxious about sharing negative feedback about their teachers, most students were able to offer both positive and critical feedback about their experiences. Although these data would be greatly strengthened with objective evidence of student learning, the rare opportunity to gather student voices about curriculum is informative and valuable.

Dissemination of Evaluation Findings

A culminating part of the evaluation involved the sharing and dissemination of the results from these innovative evaluation activities. Evaluation plans are always subject to the realities of implementation, and in this case, we adjusted our expectations so as to not burden teachers.

Of the 32 teachers who invited us into their classroom for observations and student interviews, 30 teachers provided us with students to interview for this report. We returned to the schools later and were able to interview 24 teachers about their reactions to this report. Of the six teachers from whom we did not gather data, one reported being too busy to talk, while five did not respond due to illness, departure from teaching, or unknown circumstances.

Other Efforts to Disseminate

Additional efforts were made to disseminate the results of the evaluation. These include:

- Through the Core Values Evaluation Report, Project administrators disseminated the results of the evaluation project—both the evaluation process and the evaluation of The Core Values Project—to project partners, such as the BAWP and Interactive University, and within OUSD to the Department of Curriculum and Instruction.
- The Core Values Project Evaluation is posted on the Internet.

District Response to Study Conclusions

Feedback on Report

Core Values constituents included teachers, Project administrators and the District as a whole. We sought feedback from each group during and after the report.

Teachers. Two-thirds of the teachers who we contacted had read this report in varying degrees of depth; one third had not read the report at all. For the most part, the eight teachers who did not read the report said they had not found time yet. One teacher reported not being interested in reading her students' comments. The 16 teachers who

did read the report shared a variety of responses with us. Seven teachers had a positive reaction to the report, citing parts of the report that they found helpful and interesting to them.

Three of the teachers who had a positive reaction to the report noted how students described wanting to have more choices in their assignments and research topics. One of these teachers already gives "lots of choices," and thought it was "good to see it in the report." Another teacher had "noticed that students like picking their own topic to research" after reading the report and offered students a choice of topic in a new assignment. A third teacher commented about how having a choice can help students to relate to their assignments. This teacher noted how students studying history "need connections made to their lives."

Two teachers described being eager to read the student comments. Both were curious if they could find any of their own students in the sections of quotes, and were eager to hear comments specific to their teaching. Teachers who provided four students for us to interview were able to call and talk with an evaluator who could give them specific feedback on their class, while protecting the anonymity of the individual students. Only one person called for specific comments on her classroom.

We spoke with one teacher who valued that the study had been conducted by "outsiders." This made it seem "unbiased" and gave more weight to the students' opinions. This teacher was surprised by what the students had described as boring. Before reading the student interests report, the teacher reported thinking that "the students just said something was boring because they didn't want to do it, not because it was actually more boring" than other assignments. The teacher reported that this information would help bring new awareness to what assignments could be more interesting to students.

Another teacher described reading the report and getting "something out of the comments about making activities into games." This teacher "also wants to change to show more excitement about the curriculum and more levity." Reading about how to improve "student response in class" was helpful to another teacher, also. This teacher "liked reading what teachers might do" to increase participation and involvement. Suggestions from the report might be hard to try, however, since this teacher reported being a "very traditional teacher."

In addition to the teachers who found useful information in the report, five teachers had little or no response to the Report, two teachers reported a neutral response, and another two teachers relayed a negative response to the Report. The two teachers with neutral responses explained that the report held no surprises for them, "just the usual stuff." One teacher with a negative response found the report to be "well written," but expected the report to be more about the Core Values Project and was not interested in hearing comments from the students. Another teacher with a negative response to the report found it to be "confounding in certain ways, and so generalized, it wasn't easy to follow."

Project Administrators. The Core Values Project administrators were very collegial during the data collection and report writing. They readily shared workshop attendance forms, grades, scores, attendance and other forms of institutional data that were available.

They provided introductions to principals and teachers and other necessary background information.

Although interested in the degree of implementation and impacts, after the report was completed, administrators expressed disappointment that objective measures could not show change in students' skills and knowledge.

Currently, the Technology Director is working to build a team of Core Values mentors to continue the effort. The report confirmed his own view that the Project engaged certain teachers more than others and that these teachers could become a resource.

Oakland Unified[School District. During the scope of the evaluation and Core Values tenure, many changes in the district took place. A new Superintendent and significantly different elected School Board initiated a number of changes that impacted this partnership project and that will affect similar efforts in the future, including:

- Administration and scoring of CEA tests was done on each school site, and scores were not reported to the District by teacher. A decision was made not to make the CEA scores available by teacher.
- During the partnership project, Core Values was run out of the Technology Department, much discussion took place about whether Core Values and technology integration, in general, should be the purview of the Curriculum department or the Technology department.
- Because of a decision by the new administration to return most Teachers on Special Assignment (TSA's), there was turnover among key project staff, which hurt the continuity of project efforts.

PROJECT SMART
THE OHIO VALLEY EDUCATIONAL COOPERATIVE (OVEC)
Peggy Cunningham
Gina Foletta (Department of Mathematics and Computer Science, Northern Kentucky University)

Project SMART was designed for four rural school districts to meet their needs in science, mathematics and technology as identified in each districts' Consolidated Plan. The train-the-trainer model was implemented to provide customized onsite assistance and professional development over a two-year period starting October 1999. The Ohio Valley Educational Cooperative (OVEC) used pre/post survey instruments, classroom observations, state-mandated student testing results, and inservice evaluation surveys to measure teachers' perceived knowledge in science and mathematics, comfort levels using technology as part of instruction, instructional skills, and attitudes as well as the effects of technology on their students. The analysis covering the first year and a half of Project SMART indicates some improvement in mathematics and science teaching. OVEC will finalize this evaluation of Project SMART with analysis of 2001 student test scores when they become available.

Introduction

Project SMART, "Science and Mathematics Achievement Revived through Technology," is a professional development model for educators teaching in grades K-12. This project involves sixteen schools in four rural districts in Kentucky, including eight elementary schools, four middle schools, and four high schools.

On the Comprehensive Test of Basic Skills (CTBS/5), a norm-referenced, state-mandated test, fourteen of the sixteen schools scored below the 50th percentile in 1998 in mathematics. In particular, 13 of these 14 scored below the state average, and 11 of 14 schools had declining scores from 1997 to 1998. In science, 9 of the 14 scored below the state average, and 12 of 14 had declining scores from 1997 to 1998.

Based on these data, Ohio Valley Educational Cooperative (OVEC) staff and representatives from the four Project SMART participating school districts conducted a careful analysis of the needs of each school as reflected in each of their Consolidated Plans (a school improvement plan submitted annually to the state). The most pressing needs and concerns related to professional development in science and

The Ohio Valley Educational Cooperative (OVEC) consists of 14 small rural districts in Kentucky: Anchorage, Bullitt, Carroll, Eminence, Franklin, Gallatin, Grant, Henry, Oldham, Owen, Shelby, Spencer, Trimble, and West Point. The consortium was formed to provide member school districts with the highest quality educational programs, services, and leadership through a regional approach that promotes the efficient and effective utilization of fiscal, human, and other resources. Project SMART serves four of these 14 districts; i.e., 7,100 students, most of whom are from middle to low-income families. Ninety-seven percent of students are white and 3 percent are African-American. Approximately forty-five percent of students are eligible for free or reduced-price lunch. The project includes eight elementary schools, four middle schools, four high schools, and 178 mathematics and science teachers.

mathematics identified in the Consolidated Plans were:

- lack of teacher knowledge in science and mathematics
- the need for improvement of instructional skills
- the need to align the curriculum to state curriculum documents
- the need for ongoing, authentic assessment techniques
- the need for a better teacher understanding of state and national standards
- the need for using best practices for teaching science and mathematics
- the need to integrate technology into instruction
- the need for teacher training in the content areas of science and mathematics
- the lack of teacher confidence in teaching science and mathematics and in using technology.

Project SMART was designed to meet these identified needs.

Through the train-the-trainer model, this project has affected 178 science and mathematics teachers and 7,100 students. Teachers received professional development, according to their need, in the five integrated areas of content, curriculum, instructional strategies, assessment strategies, and technology. Project staff planned for customized on-site assistance and professional development over a two-year period starting in October 1999. Staff consisted of three teachers certified in the areas of mathematics, science, and elementary education. In the Fall of 1999, project staff visited each participating school, introducing the project and calculator-based technology.

The direct professional development used technology as an avenue to provide lessons that are content specific and aligned with state and national science and mathematics standards. Each school's curriculum is based on content specified by the state curriculum standards defined in Kentucky's three curriculum documents: *Academic Expectations*, *Program of Studies*, and *Core Content for Assessment*. The science portion of the state documents includes content statements in the areas of physical, earth and space, and life science. The process skills addressed in science are scientific inquiry and applications/connections. The mathematics portion of the state documents includes content statements in the areas of number/computation, geometry/measurement, probability/statistics, and algebraic ideas. The mathematics curriculum also includes problem solving, mathematical communication, mathematics connections, and mathematical reasoning. Lessons developed by the project staff for the professional development were based on these standards and are easily incorporated into each school's curriculum. Instructional strategies used in the training modeled practices that engage students in active and meaningful learning. In addition, teachers were provided with alternative assessment activities such as portfolios, open response questions, and performance assessments that reflect real-life situations.

As shown in Table 4.1, 47 teacher-leaders each received at least 39 hours of intensive, direct professional development focusing on the integration of the five areas: content, curriculum, instructional strategies, assessment strategies, and technology.

Table 4.1
Teacher-Leader Training

Number of Teachers Trained	Spring 2000	Summer 2000	Fall 2000	Spring 2001	Summer 2001	Total
23 Elementary[1]	3 hours	30 hours	3 hours	3 hours	6 hours	45 hours
15 Middle School[2]	3 hours	30 hours	3 hours	3 hours	6 hours	45 hours
9 High School[3]	----------	30 hours	3 hours	----------	6 hours	39 hours

[1] The elementary teacher trainers delivered six to twelve hours of training to their respective staffs during the 2000-2001 school year.

[2] This direct training was available to all middle school science and mathematics teachers in the project.

[3] The high school teachers were sent to various content specific summer institutes during 2000. These teacher trainers provided cross-district workshops for the remaining science and mathematics teachers in the four high schools during the 2000-2001 school year.

During the 2000-2001 school year, 29 of these teachers attended a national science, mathematics, or technology conference to learn about current best practice in their field. Ideas and activities from these conferences were shared with colleagues.

On-site support provided by project staff addressed the assessed needs of individual schools from the five areas of content, curriculum, instructional strategies, assessment strategies, and technology. This support included modeling lessons, providing technological assistance, aiding with the alignment of curriculum to state and national documents, developing standards-based units, and delivering other customized on-site professional development. In addition, staff provided assistance through newsletters, e-mail, and phone.

Research Questions

This study addresses the following research questions:

1) Will teachers perceive improvement in their knowledge base of science and mathematics when they receive intensive professional development and on-site assistance?
2) Will teachers' comfort level in science and mathematics improve when they receive intensive professional development and on-site assistance?
3) Will teachers improve their instructional skills in the areas of science and mathematics when they receive integrated, intensive professional development and on-site assistance?
4) Will the achievement of students in science and mathematics improve when their teachers receive intensive professional development and on-site assistance?
5) How does the use of technology affect the motivation and achievement of students?
6) Will there be a difference in the changes of attitudes of teacher trainers versus non-trainers?

Methods

Table 4.2 displays a summary of the data used to answer each research question and the data collection methods.

Table 4.2
Research Questions, Data Collected, and Data Collection Methods

Research Question	Data Collected	Data Collection Methods
Will teachers perceive improvement in their knowledge base of science and mathematics when they receive intensive professional development and on-site assistance?	Teachers' perceived practices and perceived knowledge	CPE Post survey
Will teachers' comfort level in science and mathematics improve when they receive intensive professional development and on-site assistance?	Teachers' instructional practices and perceived effectiveness; teachers' comfort level in science and mathematics; teachers' instructional skills	CPE Post survey; Inservice Evaluations; classroom observations
Will teachers improve their instructional skills in science and mathematics when they receive integrated, intensive professional development and on-site assistance?	Teachers' perceived practices and perceived effectiveness; teachers' instructional skills	CPE Pre/post survey; classroom observations
Will the achievement of students in science and mathematics improve when their teachers receive intensive professional development and on-site assistance?	Student achievement	Norm-referenced, state-mandated CTBS/5 scores in mathematics; state-mandated CATS scores in science and mathematics
How does the use of technology affect the motivation and achievement of students?	Technology use and perceived effects; teachers' instructional skills; student achievement	Project SMART Teacher survey; classroom observations
Will there be a difference in the changes in attitudes of teacher trainers versus non-trainers?	Teachers' perceived practices and perceived effectiveness	CPE Post survey

The following sections describe the data sources listed as data collection methods in table 4.2:

- *The Kentucky Council on Postsecondary Education (CPE) Evaluation Questionnaire* – This pre/post survey was used to assess teachers' classroom instructional practices and perceived effectiveness as a teacher (Research questions 1, 2, 3, and 6). It was administered at the beginning of Project SMART and was given again in Spring 2001 for the purposes of a pre/post comparison of all the teachers. The final CPE evaluation was coded separately for trainers and non-trainers to facilitate making comparisons between trainers and non-trainers for the final CPE only.

- *The Project SMART Inservice Evaluation and Teacher Survey* — These instruments, designed by the project evaluator, were adapted from a previous "Teachers as Leaders" project. The *Inservice Evaluation* assessed teachers' comfort levels in science and mathematics. Teachers rated their perceived readiness to implement some of the workshop ideas (Research question 2). *The Teacher Survey* used a Likert scale (1-5) to assess teachers' technology use and the perceived effect on their students. This pre/post survey was analyzed at the school level. We had hoped to use it as a pre/post teacher survey, but we were unable to identify at the beginning of the project which teachers would be trainers (Research question 5).

- *Project SMART Classroom Observation Instrument* – This instrument was used to assess teachers' instructional skills (Research questions 2, 3, and 5). It was adapted from the Kentucky Teacher Internship Program Observation Instrument, and is based on the New Teacher Standards adopted by Kentucky's Education Professional Standards Board. These Standards compare favorably with standards and related indicators compiled by several national standards setting groups including: the National Council for Accreditation of Teacher Education (NCATE), the Interstate New Teachers Assessment and Support Consortium (INTASC), and the National Board for Professional Teaching Standards (NBPTS). The district evaluation team piloted this instrument with a seventh grade science class at Gray Middle School and an eleventh/twelfth grade trigonometry class at Ryle High School in October 2000. Ratings for the most part were similar. On those items where there was discrepancy, the team revised the instrument. The observation protocol has three major components: knowledge of content, instructional implementation (e.g., inquiry based, use of technology, cooperative group work), and types of assessments.

- *Comprehensive Test of Basic Skills (CTBS/5)* – This norm-referenced test is state-mandated for grades three, six, and nine. We collected mathematics teachers' student data for 1999, 2000, and 2001. These data were used to assess improvement in student achievement (Research questions 4 and 5).

- *Commonwealth Accountability Testing System (CATS)* — Test results in mathematics and science for 1999 and 2000 were used to assess improvement in student achievement (Research questions 4 and 5). The Kentucky designed and state-mandated CATS tests are

not numerically scored. Students are rated as novice, or apprentice, or proficient, or distinguished.

Successes and Challenges

Difficulties Using Student Test Score Data

We have CTBS/5 data from 1998 through 2001 for mathematics only since the Kentucky Department of Education did not select science for standardized testing. We have 1999 and 2000 CATS data for both mathematics and science. Because the 2001 CATS scores will not be available until September, we could not use them in this analysis. In mathematics, the CATS test is given at the end of grades 5, 8, and 11. The only science data we could use is from CATS, given at the end of grades 4, 7, and 11. Problems arise when we have a lead science teacher who teaches grade 6, thus not having student data available. Therefore, we used CATS data for five lead science teachers at grades 4 and 7; and CATS or CTBS/5 data for 11 mathematics teachers at grades 3, 5, 6, and 8. We did not include the 1998 KIRIS (a previous Kentucky state-mandated test) scores because they are not comparable to the CATS scores at this time.

The task of looking at these teachers' student gains was a challenge. In Kentucky, students in grades 4 through 12 are required to take the CATS Test. The data is reported to the districts by school and students rather than by teacher. We searched teacher and student records to obtain the information we need. Another challenge was that over the three-year period that we looked at the CATS data for each of our teachers and their students, we looked at different students each year. In the final analysis, we were unable to obtain teachers' students data, but only school data for specific grade levels. In cases where there was only one science or mathematics teacher at a particular grade level in the school, the school score does reflect upon the teacher.

Difficulties with Fluidity of Teachers' Participation in Project

The CPE Evaluation Questionnaire was administered to 178 teachers in summer 1999 and 123 teachers in June 2001. The pre-questionnaire does not separate out the lead teachers from the rest of the teachers because we did not know who our lead teachers would be; whereas, we did distinguish between the lead teacher and the rest of the teachers on the post-questionnaire. Some of the teachers had left the program while a few did not follow through by completing the survey.

We have Inservice Evaluations from different time periods: 25 elementary teacher-trainers in January 2000, 16 middle school teachers in February 2000, 32 middle school teachers and 30 elementary teachers in summer 2000 (The latter two numbers reflect some non-Project SMART teachers. They have been coded so that we can remove them from the pool when we do our final analysis), 24 middle school teachers in September 2000, 18 elementary teachers in October 2000, 16 middle school teachers in February 2001, 25 elementary teachers in March 2001, 26 elementary teachers, and 19 middle school teachers in June 2001 and eight high school teachers in July 2001.

We have teacher surveys from 15 middle school teachers' February 2000 workshop. They completed the same survey at the end of the project in June 2001. This enabled us to do a pre/post analysis at the school level for the middle school teachers.

Difficulties and Successes with Classroom Observations

The project staff conducted 32 classroom observations for 26 of the 47 teacher-leaders – 18 observations in elementary schools, nine observations in middle schools, and five observations in high schools. Each of five elementary teachers and one high school teacher were observed twice. Those teachers who were observed more than once had consistent ratings among observers. We wanted to observe each teacher-leader more than once, but there was insufficient time and staff to do it. It was important for us to observe each lead teacher who stayed with the project at least once, and then follow-up with additional observations when possible. We observed only 3 of the high school teachers' classrooms. The remaining 6 were observed while they provided training for their colleagues. The teachers were very cooperative and welcoming as we spent time in their classrooms throughout the year. It was especially rewarding to observe them and their students effectively using the technology provided by the project.

Findings

Each of the following sections pertain to one of the six research questions stated in Table 4.2. We will discuss the findings in order of each research question.

Teachers' Perception of Improvement in Science and Mathematics Knowledge Base

All of the teachers affected by Project SMART expressed some degree of improvement in their knowledge base and instructional practices. Using a Likert scale from 1 to 5 where 1 means strongly agree, 3 is neutral, and 5 means strongly disagree, the mean responses ranged from 1.9 to 2.7 as displayed in Table 4.3. As anticipated the teacher trainers felt more strongly about gains in knowledge and instructional approaches than the non-trainers.

Table 4.3
CPE Post Survey Results on Selected Items

Item	Mean for All Teachers	Mean for Trainers	Mean for Non-Trainers
I learned new concepts, facts, and definitions	1.9	1.4	2.0
I have a better understanding of fundamental core content in my discipline	2.5	1.8	2.7
I learned new instructional approaches	1.9	1.2	2.0

Teachers' Comfort Level in Science and Mathematics

The Inservice Evaluations indicate that most of the teacher-leaders are gaining information enabling them to implement in their classrooms some of the ideas presented in the Project SMART training. Table 4.4 below displays the number of teachers responding "Yes" (from these choices: Yes, No or Partially) to the OVEC evaluation item: "Do you feel you learned enough to begin using some of these ideas and materials in your own teaching?" For the item from the T^3 workshop – "This institute helped make me feel more comfortable about using handheld technology in my classroom" – the teachers responded to a Likert scale: Agree strongly, agree somewhat, neither agree nor disagree, disagree somewhat, disagree strongly. Most entries in this category were "Yes," with the others representing summaries of Likert-type responses to the T^3 Workshop.

Table 4.4
Teachers Reported Readiness to Implement Reform

Time	Training Type	Number of Teachers
Spring 2000	January OVEC 6-hour Workshop	20 of 25 elementary
	February OVEC 3-hour Workshop	11 of 16 middle school mathematics and science
Summer 2000	June 30-hour technology institute by OVEC	31of 32 middle school mathematics and science
	June T^3 workshop	24 of 30 elementary agree strongly 6 of 30 agree somewhat
Fall 2000	September OVEC 3-hour Workshop	18 of 19 middle school mathematics and science
	October OVEC 3-hour Workshop	14 of 18 elementary
Spring 2001	February OVEC 3-hour Workshop	16 of 16 middle school mathematics and science
	March OVEC 3-hour Workshop	13 of 13 elementary trainers; 8 of 12 non-trainers
Summer 2001	June OVEC Workshop	26 of 26 elementary teachers; 19 of 19 middle school teachers
	July OVEC 6-hour Workshop	7 of 8 high school teachers

Of the 178 teachers affected by Project SMART, only the 26 of the 47 teacher-leaders were observed. These classroom observations support the teacher-leaders' responses on the Inservice Evaluations. The four observers rated the teachers as demonstrating one of the following: satisfactory progress, improvement needed, not satisfactory, or not applicable. For the "uses technology with confidence" item on the instrument, all but eight of the 26 teacher trainers were rated at "satisfactory progress". These eight teachers were rated

"improvement needed" on the first observation and six of these eight were rated at "satisfactory progress" on subsequent observations. The remaining two teachers were not observed a second time due to time restraints. All teachers were rated as at "satisfactory progress" on all other items on the instrument except for three teachers who were rated as "improvement needed" on inquiry related instruction and two teachers who were rated "improvement needed" on "makes creative and appropriate use of the calculator/CBL

These responses from the Inservice Evaluations and classroom observations are supported by the teachers' responses on the CPE Post survey (see Table 4.5) given after almost two years of customized onsite assistance and professional development as part of their participation in project SMART. Again, the non-trainers' comfort level in teaching was not as strong as for the trainers. Using a Likert scale from 1 to 5 where 1 means strongly agree, 3 is neutral, and 5 means strongly disagree, the mean responses ranged from 1.5 to 2.5.

Table 4.5
CPE Post Survey Results on Selected Items

Item	Mean for All Teachers	Mean for Trainers	Mean for Non-Trainers
I am an effective teacher	2.2	1.5	2.4
I am more excited about teaching in my subject area	2.3	1.6	2.5

Teachers' Improvement in Instructional Skills in Science and Mathematics

Using the CPE Questionnaire, the teachers responded to the 7 item pairs shown in Table 4.6 at the beginning of the project in October 1999 and near the end of the project in May 2001. Each pair represents opposite ends of a continuum to approaches in classroom teaching. The teachers were instructed to circle the number that best describes their position on the continuum. In comparing the pre/post means for each of the item pairs A through G, no significant differences were found.

Table 4.6
CPE Pre/Post Items

	1	2	3	4	5
A	Classroom interaction consists of teacher-led lecture with limited response from students				Classroom interaction involves a dialogue among teacher and students
B	Students generally work in groups cooperatively				Students generally work independently
C	Instruction focuses on the central ideas of a discipline, covering less topics in depth				Instructional emphasis on broad coverage of information with little depth
D	Student role is to receive/recite factual information or employ rules and algorithms through repetitive routines				Student role is to manipulate information and ideas in ways that transform their meaning and implications
E	Students generally learn concepts and processes utilizing hands-on approaches				Students generally learn concepts and processes involving readings, lectures, and demonstrations
F	I am generally successful in encouraging effort and participation among all students				I find it difficult to encourage the efforts and contributions of certain students or groups of students
G	I generally assess students' progress using conventional methods (e.g., paper and pencil tests such as multiple choice, fill-in-the-blank, true/false)				I generally assess students' progress using alternative methods (e.g., open-response questions, hands-on performance, portfolios, observations)

Pair	Pre Mean_All	Post Mean_All
A	4.18	4.17
B	2.87	2.72
C	2.65	2.62
D	3.67	3.81
E	2.55	2.30
F	2.07	2.08
G	3.42	3.42

During classroom observations, we looked for the following behavior as indicative of classroom instruction and inquiry teaching:

- Includes learning experiences that encourage students to be creative, resourceful, flexible, and adaptable
- Motivates, encourages, and supports individual and group inquiry
- Elicits samples of student thinking and stimulates student reflection on their own ideas
- Effectively applies methods of inquiry related to certified academic areas

These items were rated as satisfactory progress for all but three of the 26 teacher-leaders. Of the three rated lower, two received "improvement needed" and one received "not satisfactory."

Student Achievement in Science and Mathematics

The CTBS/5 schools' mathematics percentile scores shown in Table 4.7 indicate significant overall improvement from 1999 to 2001 at the $p = .02$ level, using ANOVA simple factor analysis. The district scores at the elementary level for County C are included in both tables below. In other counties the test was given at only one elementary, middle school, and high school in each district; thus, those schools' scores represent the districts' scores. These results cannot be credited solely to the project since different students are tested each year.

Table 4.7
Students' CTBS/5 Scores Over Time

School	Subject	# Project Teachers	# Teachers Total	1999	2000	2001
County A High	Math-9th	1	4	39	42	50
County B Elem.	Math-3rd	2	2	45	38	67
County B Middle	Math-6th	1	1	42	36	44
County C Elem.1	Math-3rd	2	2	57	42	65
County C Elem.2	Math-3rd	1	4	41	47	54
District	Math-3rd			48	44	
County C Middle	Math-6th	1	1	53	46	51
County C High	Math-9th	1	4	47	48	55
County D Middle	Math-6th	1	1	44	56	48.7
County D High	Math-9th	1	3	47	57	56
State Summary	Math-3rd			51	55	
State Summary	Math-6th			49	50	
State Summary	Math-9th			46	47	

Our interim mathematics and science CATS school data shown in Table 4.8 indicate no significant improvement from 1999 to 2000. However, the 2000 test was given six months

after the onset of Project SMART. The range for the CATS test is from 0 to 140, with 100 being proficient. A score of 100 for every school by the year 2014 is Kentucky's state goal.

Table 4.8
Students' CATS Scores Over Time

School	Subject/ grade	# Project Teachers	#Teachers Total	1999	2000
County A Elem.	science/4	1	3	47.6	52.0
	math/5	1	3	63.0	68.1
County A High	science/11	2	4	61.0	61.1
	math/11	1	4	64.4	58.9
County B Elem.	science/4	1	2	48.5	46.3
	math/5	1	1	45.5	49.7
County B Middle	science/7	1	2	32.6	34.7
	math/8	1	1	50.3	57.3
County B High	science/11	3	3	66.1	67.4
County C Elem.3	science/4	1	2	49.8	52.9
County C Elem.2	science/4	1	1	62.2	64.7
	math/5	1	1	60.7	66.5
District	science/4			57.9	57.7
	math/5			58.9	63.8
County C Middle	math/8	1	1	66.2	72.7
County C High	science/11	1	4	61.9	58.7
	math/11	2	5	66.6	61.1
County D Elem.	math/5	1	6	65.6	64.2
County D Middle	science/7	1	1	38.3	34.9
	math/8	1	1	76.5	76.1
County D High	science/11	1	3	72.6	74.1
	math/11	1	3	72.3	81.3
State Summary	science/4			55.1	56.9
	math/5			64.0	67.1
	science/7			36.8	37.4
	math/8			67.0	70.6
	science/11			66.3	67.1
	math/11			67.2	68.3

Technology Effects on Motivation and Achievement of Students

The Project SMART Teacher Survey was given to 31 elementary and middle school teacher-trainers. Selected items from this survey give us insights into the teachers' perceived effects of using technology with their students. Using a Likert scale from 1 to 5 where 1 means

strongly agree, 3 is neutral, and 5 means strongly disagree, the responses ranged from 1.5 to 2.1 as displayed in Table 4.4.

Table 4.9
Project SMART Teacher Survey Results on Selected Items

If you have used technology to teach, please answer the following:	Mean (1-5)
Most of my students enjoy using technology to learn.	1.5
Most of my students seem to learn better when we use technology.	2.1
I have some students who are generally apathetic and unsuccessful when it comes to learning yet they get involved and are successful when we use technology.	2.0

The teachers agree that their students enjoy using technology and tend to learn better when using technology.

Since the use of technology was an emphasis for this project, we realize that its use or non-use can affect students' motivation and subsequently, achievement. As we observed teachers in the classroom, we found that effectively using technology was the most challenging project component for them.

During these observations, we looked for the following behavior as indicative of technology use in the classroom:

- Includes creative and appropriate uses of calculator/CBL (Calculator Based Laboratory)
- Makes creative and appropriate use of calculator/CBL
- Utilizes calculator/CBL related to academic areas
- Uses technology for individual, small group, and large group learning activities
- Uses calculator/CBL to support multiple assessments of student learning
- Uses technology with confidence

These items were rated as satisfactory progress for all but eight of the 26 teacher-leaders. As noted earlier, these eight teachers were rated "improvement needed" on the first observation and six of these eight were rated "satisfactory progress" on subsequent observations. The remaining two teachers were not observed a second time due to time restraints.

Attitudes of Trainers versus Non-Trainers

Using the CPE Post survey, the teachers responded to the seven item pairs (shown in Table 4.6) near the end of the project in May 2001. Each pair represents opposite ends of a continuum to approaches in classroom teaching. The teachers were instructed to circle the number that best describes their position on the continuum. In comparing the Table 4.10 post means of trainers versus non-trainers for each of the item pairs A through G, no significant differences were found.

Table 4.10
Perceived Instructional Practices for Non-Trainers versus Trainers

Pair	Mean Non-Trainer	Mean Trainer
A	4.18	4.11
B	2.74	2.61
C	2.60	2.72
D	3.81	3.83
E	2.30	2.29
F	2.12	1.94
G	3.47	3.22

Conclusions

This preliminary analysis seems to indicate positive trends for both the teachers and their students. The CPE survey supports that all of the teachers affected by Project SMART expressed some degree of improvement in their knowledge base and instructional practices. The teacher trainers felt more strongly about gains in knowledge and instructional approaches than the non-trainers. (Research Question 1) Our classroom observations and Inservice Evaluations support our preliminary conclusion that the teachers' comfort level and instructional skills in science and mathematics improved (Research Questions 2 & 3). The CTBS/5 data show improvement in schools' student achievement in mathematics over time. (Research Question 4). In addition the interim CATS results show some increase in student achievement (Research Question 4). The Project SMART teacher survey and classroom observation findings indicate that their students enjoy using technology and tend to learn better when using technology (Research Question 5). The CPE Post survey reflects no differences in attitudes of trainers and non-trainers (Research Question 6).

LITERACY TRAINING
PLAINFIELD PUBLIC SCHOOLS
Michael Wilson

Plainfield Public Schools used Goals 2000 funds to support a variety of professional development to K-8 teachers in language arts. Upon completion of the professional development program, teachers are expected to conduct language arts instruction in a two- to two-and-a-half hour block (depending on grade level) that emphasizes individual and small group work. The district evaluation collected data via student interviews, teacher surveys, district records and databases, as well as state test scores in reading and writing. The district sought to ascertain whether the language arts training had been implemented, and was working, as designed; and what effects the implementation had had on student achievement in language arts. Although no statistically significant outcomes resulted in this study, the program is being implemented at many schools and future evaluation — with improved data collection methods — is strongly recommended.

Introduction

In 1999-2000, state-administered tests revealed that Plainfield's Elementary School Proficiency Assessment (ESPA) language arts scores were slightly higher than the year before, but in language arts sections of the Grade Eight Proficiency Assessment (GEPA) and the High School Proficiency Test (HSPT11), the scores were relatively unchanged. More importantly, all of these scores were lower than the average state scores.[1]

Two years earlier, the school board decided that the most pressing need in Plainfield Public Schools was to ensure the students mastered grade-level skills in language arts, which in the New Jersey Core Curriculum Content Standards include reading, writing, listening, speaking, and viewing. For the next two years, the school district focused on improving learning in these areas in the training that was accomplished as well as the reform efforts that were accomplished in each of the schools.

Plainfield Public Schools is in a community of 48,000 inhabitants of low to middle income families. The school district, with over 7,000 students, has ten elementary schools, two middle schools, one high school, and one adult school. The school district employs about 900 full-time staff including instructional, non-instructional and administrative personnel. About 71 percent of the students attending Plainfield Public Schools are eligible for free and reduced lunch. The students in Plainfield are about 81 percent African-American, 17 percent Hispanic, and 2 percent other.

[1] In the language arts section of the ESPA, 56.5% were partially proficient, 43.1% were proficient, and .4% showed advance proficiency. In Mathematics, 58.1% were partially proficient, 37.3% were proficient and 4.6% showed advance proficiency. In the language arts section of the Grade Eight Proficiency Assessment (GEPA), 52.8% were partially proficient, 47.2% were proficient, and 0.0% showed advanced proficiency. In mathematics, 77.2% were partially proficient, 21.9% were proficient, and 0.9% showed advanced proficiency. Finally in the October, 1999 Eleventh Grade High School Proficiency Test (HSPT11), 65.3% of the students passed the reading section, 65.1% passed the mathematics section, and 71.2% passed the writing section of the test.

Research Questions

This study addressed the following research questions:

1) Is training in language arts being implemented as designed?
2) Is the coaching component in language arts working as designed?
3) Does coaching have an effect on the implementation process?
4) Does the implementation of the elements of language arts training have a positive effect on the achievement of student in language arts?

Description of the Language Arts Program

In Plainfield Public Schools, a basic element of the language arts program is the "literacy block." Designed by America's Choice, the literacy block is a whole-school reform model that the state requires Plainfield schools to implement to support its academic program. The literacy block is an organizational tool used by classroom teachers to engage students in a variety of literacy experiences. During the block, students participate in whole group, small group and individual activities related to reading, writing, speaking, listening, and viewing. In grades K-3, the literacy block is two-and-a-half hours long. In grades 4-5, the literacy block is two or two-and-a-half hours long. Decisions concerning instruction are based on ongoing teacher assessment of student strengths and weaknesses (e.g., through running records, observations of students, conferencing with students, review of student written work, teacher-made assessments, and the Target Assessment Process).

Two required components of the literacy block include the Reading Workshop and the Writing Workshop. These components are linked together in two ways: (1) through the oral language that surrounds, supports, and extends all activities; and (2) by the content or topic of focus.

- Reading Workshop
  This is an individualized approach to reading designed to assess and assist each student's reading progress. During Reading Workshop, students self-select developmentally appropriate reading material under the guidance of the teacher. Students read independently, keep a response journal and reading log. They complete a variety of assignments designed to assess comprehension, and they conference regularly with their teacher. Reading Workshop allows students to independently practice skills, concepts, and strategies using authentic literature.

- Writing Workshop
  This is structured time for teaching and learning about writing. It includes time for a whole group mini-lesson based on what most students need to learn about writing. Children spend a good portion of the workshop time working on their own writing while the teacher confers with individuals or a small group. At the end of the workshop, there is a short sharing time for writers to give feedback to each other.

The teacher spends a minimum of 30 minutes each day in developing reading/writing skills through direct instruction using the core grade-level textbook and/or other grade-level literature.

The skills block is required for grades K-3. About 90 percent of the grades 4-5 had the same or more time for literature, but the program was only required for the K-3 level classes.

A 60-minute language arts block could include different aspects of either Reading or Writing Workshop and could vary from day to day.

Plainfield has established a training program to meet the needs of teachers and administrators in the district. The following describes the training program's goals, how the training is aligned with state and district curriculum content and standards, the type of activities used in the training program, and how the America's Choice reform program is part of the training.

District training program goals

To provide effective staff training that:

- Requires, motivates and supports the norm of continuous improvement which provides changes that sustain and accelerates movement towards district and state goals associated with academic excellence and life-long learning.
- Engenders and supports the increase of knowledge, skills, processes and attitudes regarding organizational and system thinking, human learning and development, effective use of student data, instructional effectiveness, and increases in cognitive achievement, and goal directed learning.

Alignment of training with state/district curriculum content and standards

Nearly all aspects of the training process in Plainfield Public Schools are directed to supporting the New Jersey Core Curriculum Content Standards.

- Most of the program itself is broken into important elements of the standards, i.e., Literacy Language Arts, Mathematics, Science, Technology, and World Languages.
- The organization of the program is based upon the Training Standards, one aspect of the State's complement of standards.
- The district curricula include Language Arts, Mathematics, Science, Social Studies, and Fine and Performing Arts. These curricula are based on the New Jersey Core Curriculum Content Standards and provide the basis for the instruction in most of the teacher training activities in Plainfield.
- All teachers' lesson plans must include a reference to the standards.

Type of activities

The following strategies for professional learning represents the range of staff development practices used in the Plainfield Public Schools Training Program.

- **Coaching and Mentoring** – This refers to working one-on-one with an equally or more experienced teacher or other professional through activities such as classroom observation and feedback, problem solving and troubleshooting, joint planning, strategy implementation, and discussion of classroom processes. This method of training has become particularly prominent since it is an important element in the America's Choice Whole School Reform to which most of the schools in Plainfield are subscribing. The schools using this model are required to hire design coaches and literacy coaches, teachers who spend the major part of the time involved in coaching.
- **Developing professional developers** – This approach focuses on building the skills and knowledge to create learning experiences for other educators in selected staff. These professional developers provide on-site assistance by presenting, demonstrating, and supporting teacher development.
- **Professional networks** – In these networks, teachers are able to explore and discuss topics of interest, share information and strategies, and identify and address common problems with other teachers.
- **Reflection on student work and student thinking through assessment and evaluation** – Through the careful examination of student work and products, teachers are able to develop powerful reflections on not only the development of student academic achievement and developmental progress but also their own instructional program and training needs.
- **Study groups** – By engaging in regular, structured, and collaborative study and discussion of topics identified by the group, staff members have the opportunity to examine new information, reflect on classroom processes, and become part of a learning community.
- **Case discussion** – There are increasing numbers of written narrative or videotapes of classroom teaching and learning available that can provide the basis for rich staff development through discussion and problem solving activities.
- **Action Research** – Engaging in a research project in the classroom allows teachers to examine their own teaching and their students' learning.
- **Workshop, institutes, courses, and seminars** – Structured opportunities outside the classroom allow staff to focus intensely on topics such as the content of mathematics, science, or English literature and to learn from others with more expertise in these areas. To be effective, these experiences are integrated within the context of regular classroom instruction using other techniques described above.
- **Curriculum development adaptation** – Creating and adapting new instructional materials in collaboration with others provides growth to teachers and allows teachers to go beyond the confines of their own classroom.
- **Graduate coursework** – Continued graduate work by staff is important to creating a learning community in Plainfield.
- **Clinical experiences** – This strategy occurs when staff are involved in the district application of new learning and then become part of an effort to reflect on the results of the use of that new learning. It is essentially professional development in the act of teaching.

America's Choice training component

One of the key components of the America's Choice whole-school reform program is coaching. Each school program includes a literacy coach and a design coach. Both coaches help train teachers to function in the literacy block.

The literacy coach is required to implement various training activities, including: demonstration teaching, one-on-one observation and teacher coaching, and facilitating group observations and critiquing. To accomplish the demonstration teaching, the literacy coach visits different teachers' classrooms and teaches specific lessons to the teacher's students, pointing out the different aspects of the lesson that the teacher needs to follow. Once a teacher has observed a demonstration lesson, the literacy coach observes the teacher in a one-on-one situation as she or he uses the same techniques that were used in the demonstration lesson. After the lesson, the literacy coach gives the teacher coaching feedback on the different elements of the lesson format as the teacher's lesson related to the original demonstration lesson. The literacy coach can conduct this same observational technique in a group situation in which several teachers observe both the demonstration lesson and the teacher's subsequent attempts to replicate the elements of the demonstration lesson.

Literacy coaches also arrange additional time and instruction in tutorial programs provided by classroom teachers for students whose achievement is less than other students in their class. The literacy coaches use the techniques described above to develop the tutoring skills of the teachers. Finally, the literacy coaches are responsible for implementing an assessment program that allows them to determine whether the students of participating teachers are making anticipated progress.

The design coach also works in the area of literacy. However, design coaches' tasks differ from literacy coaches. The primary task of the design coach is to establish a model classroom. In this model environment, the design coach provides groups of teachers with the processes that support the literacy block, including specific types of books, charts, furniture arrangement, data recording devices, as well as appropriate instructional strategies. The design coach helps teachers use the design they observe in the model classroom by interacting with them during lessons and by observing and critiquing their use of the strategies in the model classroom. In addition to this work in the model classroom, design coaches also provide professional development to groups of teachers and assist teachers in developing school, classroom, and individual student plans, analyzing student performance data, and relating state standards to instructional outcomes.

Methods

Table 5.1 below summarizes Plainfield's study questions and the data sources and methods used to attempt to answer those questions.

Table 5.1
Research Questions, Data Collected, and Data Collection Methods

Research Question	Data Collected	Data Collection Methods
Is training in language arts being implemented as designed?	Students' self-report regarding implementation of reading and writing blocks	Implementation Interview Protocol
Is coaching in language arts working as designed?	Teachers' self-report regarding instructional practices	Language Arts Literacy Survey and Coaching Questionnaire
Does coaching have an effect on the implementation process?	Teachers' self-report regarding instructional practices and students' self-report regarding implementation	Language Arts Literacy Survey and Implementation Interview Protocol
Does the implementation of the elements of language arts training have a positive effect on the achievement of student in language arts?	Student test scores; student and teacher self-reports regarding implementation	TAP writing, GRADES 1-5; Goals Performance Assessment in reading Grades 1-5; ESPA, GRADE 4; Language Arts Literacy Survey and Implementation Interview Protocol

Data sources included the following:

- *District databases* – These databases include objective achievement data for students currently enrolled with about four years of testing data. For those students who are not new to the district, the database includes state testing data, district testing data (on the Goals Performance Assessment) and data from an instrument administered three times a year called the Target Assessment Process (TAP).

  There is limited data on teachers. It is, in fact difficult to determine what teachers assignments are and how long they have been teaching in the district without directly contacting the school or asking for the information from the Human Resources Department.

- *Coaching questionnaire* – A questionnaire was used to collect information from the literacy coaches. The questionnaire was generated using the objectives for the literacy coaching program and the job description established for the literacy coaches. The questionnaire was designed around a Likert response format. This questionnaire provides information on the extent to which coaching is proceeding as designed. Teachers who work with the coaches completed the questionnaires.

- *Implementation interview protocol*– An interview protocol was generated to gauge the level at which each classroom was involved with the reading and writing workshop components of the language arts program. The questionnaire was generated using objectives and description of the reading and writing workshops. It was administered to a group of students from each classroom in each school to determine to what extent the students understood and were involved in the reading and writing workshops.

  In 1999-2000 school year, district administrative personnel observed the implementation of important elements of the language arts curriculum in each classroom in the district's schools. In the original plan of the evaluation, these observations were to be used as data showing the level of implementation of the language arts (reading and writing) curriculum in Plainfield Public Schools. After a closer analysis of these data it was determined that there were serious problems with consistency of construct interpretation among observers, and it was decided not to use these results. Instead of the observational data, the level of implementation was measured indirectly using the protocol described above.

- *Goals Performance Assessment* – The Goals Performance Assessment instrument was developed by Harcourt Brace. It is an instrument that is completely based on students' written responses to questions. There are two examinations, one in mathematics and one in reading. The instrument is approximately 10 items long in each of the subjects tested. The analysis of the instrument provides both norm-reference scores including percentile rank, scale scores, and normal curve equivalent scores as well as performance indicators. The instrument is reader scored using a rubric with a scale of 0 to 3 for each of the ten questions, 0 being no mastery, 1 partial mastery, 2 mastery, and 3 advanced mastery. This instrument was chosen because the content and response mode reflect important aspects of the New Jersey Core Curriculum Content Standards and state assessments. By the spring of 2000, four years of data had been collected using the Goals Performance Assessment. These data were examined longitudinally to determine to what extent the implementation of language arts training and use of coaching may have had an impact on reading scores.

- *Target Assessment Process (TAP) in reading and writing* – The TAP is a district-developed instrument that provides teachers and student with a general idea of what is expected in each grade, referred to as the target, and data that show what kind of progress students are making toward the target. TAP writing is a single essay prompt and TAP reading consists of three to five reading comprehension questions depending on the grade level of the reader. The reading and writing TAPs are based on the district curriculum which is in turn is based on state standards. These curricula provide grade by grade breakdowns of student achievement outcome expectations. TAP is given three times a year. The end-of-year TAP results were used to determine to what extent language arts achievement relates to implementation and coaching. Since the TAP instrument is based on the state standards, it was used to clarify the relation between the elements of the language arts training, all of which are standards-based.

- *Elementary School Proficiency Assessment (ESPA)* – This instrument was developed and scored by the State Department of Education. The language arts section tests reading, writing, and viewing and provides an overall language arts score. The language art section is composed of multiple choice and constructed response questions.

- *The Language Arts Literacy Survey* – This instrument was developed strictly for this study. It has not been used before nor will it be used after the study has been completed. It contains a series of questions to understand to what extent each teacher is involved in the coaching related language arts literacy program development training offerings and how many of the language arts literacy initiatives art being used by the teacher. This instrument was developed after it became obvious the coaching and training data originally planned for use were too flawed from inconsistent interpretation and recording variations to be useful.

Findings

The evaluation focused specifically on the language arts training and the effectiveness of the use of coaches in the implementation of that training. During the evaluation, data were gathered about the level of implementation of the different aspects of the language arts program which included the various elements of Reading and Writing Workshop and the training and coaching and used to support these two. Of particular interest was whether and to what extent to literacy and design coaches were effectively involved in the implementation of the language arts processes. These data along with achievement data were used in a structural linear equation model to determine the extent to which the different aspects of the language arts training program are functioning as designed. A second analysis used longitudinal data in a mixed effect-blocking model.[2] This analysis addressed the effect of teachers at various levels of training using four years of norm-referenced achievement data from an instrument aligned with state standards.

Question One: Is training in language arts being implemented as designed?

This question is answered using the results to the Interview Implementation Protocol given to groups of students from each classroom.

Table 5.2 shows the responses of students in each classroom to questions on the Interview Implementation Protocol about the implementation of the reading and writing blocks in their respective classrooms. The table provides the number of questions answered correctly out of the total number of questions. The rate of reading block implementation shows a low of 18 percent at Stillman Elementary School and a high of 65 percent at Cedarbrook Elementary School. The survey was an attempt to obtain information on each teacher on the level of implementation of the writing and reading blocks by asking students about what was occurring in the class room and to gain some understanding of their understanding of the processes and the kinds of learning they derived from their experiences with the programs. For each question that the students were able to answer, the response to the survey question was interpreted as an indication that the teacher was implementing that aspect of the program well enough for the students to understand and benefit from the teacher's implementation process. A teacher with a score of 50 percent would indicate that the teacher's students answered 50 percent of the questions that the interviewer asked of the students. All of the teacher's ratings were averaged and the average rate has been reported for the school. The average rate of implementation of the writing block shows a low of 27 percent at Stillman and Clinton and a high of 38 percent at Cedarbrook. Comparing these average rates across the schools shows less variation across schools in the rate of implementation in the writing block than in the reading block, such that the average rate of

[2] This model was described by William Sanders from the University of Tennessee.

implementation in the writing block is lower than that of the reading block in 7 of the 10 elementary schools.

Table 5.2
Percent of Correct Answers for Reading and Writing Blocks on the Interview Implementation Protocol by School

School	Reading Implementation	Writing Implementation
Barlow	20%	28%
Cedarbrook	65%	38%
Clinton	41%	27%
Cook	50%	36%
Emerson	40%	32%
Evergreen	32%	36%
Jefferson	50%	35%
Stillman	18%	27%
Washington	53%	36%
Woodland	36%	32%

Continuing the pattern, the level of reading implementation seems to be higher than writing at all grades except fifth grade (see Table 5.3). Both the reading and the writing implementation appear to be lowest at the lower grades but in no case do either appear to be over 50 percent. Hence, it appears that implementation is unevenly developed across both grades and schools and has considerable room to grow for a relatively complete implementation.

Table 5.3
Percent of Correct Answers for Reading and Writing Blocks of Language Arts Literacy Survey, by Grade

Grade	Reading Implementation	Writing Implementation
First Grade	30%	17%
Second Grade	39%	29%
Third Grade	46%	35%
Fourth Grade	48%	44%
Fifth Grade	41%	44%

Question Two: Is coaching in language arts working as designed?

Coaching frequency shows a percent of teachers who answered in each of the five categories of questions on the Language Arts Literacy Survey. The question about frequency of coaching asked teachers "how frequently do you interact with, observe something presented by someone, or are observed teaching something by someone, either a literacy coach or Curriculum and Instruction staff member, on writing or reading workshop. In the survey teachers could respond: At least once a day, at least once a week, at least once a month, less than once a month or never. Coaching Experience refers to the question in the Language Arts Literacy Survey (question #2) relating to the number of ways that teacher experienced different aspects of coaching. Coaching Use refers to the next question (#3) related to the number of different ways that coaching is used or implemented in the classroom. Tables 5.4 and 5.5 show results of the three main sections of the coaching questionnaire that was sent to teachers to complete. Table 5.4 shows a breakdown of the percent of teacher responding by category for each school and Table 5.5 shows a percent by grade. The first section in each table, Coaching Frequency, shows the percent of teachers who responded in each category. The next section in the chart shows the average across all of the teachers for coaching experience and use.

The information in the tables shows that the highest frequency of coaching was experienced at Cook, Clinton, Evergreen and Woodland Elementary Schools, and the lowest frequency was experienced at Barlow, Jefferson, Stillman and Washington. In terms of grades, the greatest frequency was experienced at the lower grades and the lowest frequency at the higher grades. As with implementation, the variation is considerable from school to school and grade to grade. In six of the schools, more than fifty percent of the teachers reported receiving coaching an average of only once a month or less. In four of the schools, teachers on average received coaching an average of more than once a month.

In the second part of the two charts, the number of types of coaching experienced by each teacher was fairly low. Teachers were asked to choose from up to eight different types of literacy-related coaching experiences in which they had participated, such as observing a model classroom, participating in a Core Assignment, or receiving critique from a literacy coach. Two of the choices were open-ended. For five of the schools (Cedarbrook, Clinton, Cook, Washington and Woodland) teachers reported that they experienced an average of more than two of the eight possible experiences. On average, the teachers at the rest of the schools reported fewer than two experiences.

Teachers also reported on points at which they used what they had learned in coaching. Teachers were given a list of six instances in instruction (such as Writing Workshop, Reading Workshop, and lesson planning) during which they might use information from their coaching sessions. The result show that teachers in seven of the ten schools used on average about three of the possible six instances available to teachers. In four of the five grades, teachers experienced close to three of the possible uses. Therefore, while the experiences of coaching seem restricted, the possible uses of coaching appear to be slightly more varied.

Table 5.4
Coaching Frequency, Coaching Experience, and Coaching Use, by School

School	Coaching Frequency					Coaching Experience	Coaching Use
	Never	<1 session /month	>1 session /month	>1 session / week	>1 session / day	total possible = 8	total possible = 6
Barlow	13%	75%	13%			1.1	2.1
Cedarbrook		23%	31%	46%		2.9	3.2
Clinton	18%		18%	46%	18%	2.5	2.5
Cook			36%	36%	27%	2.2	3.1
Emerson		53%	11%	32%	5%	1.6	2.9
Evergreen	6%		33%	56%	6%	1.9	2.7
Jefferson	25%	75%				1.2	2.4
Stillman	13%	88%				1.0	1.5
Washington	33%	17%	17%	33%		2.0	2.0
Woodland		9%	9%	73%	9%	2.6	3.0
District	9%	32%	18%	35%	7%	1.9	2.6

Table 5.5
Coaching Frequency, Coaching Experience, and Coaching Use, by Grade

Grade	Coaching Frequency					Coaching Experience	Coaching Use
	Never	<1 session /month	>1 session /month	>1 session / week	>1 session / day	total possible = 8	total possible = 6
First Grade	4%	28%	12%	48%	8%	2.8	2.9
Second Grade	13%	29%	8%	38%	13%	2.1	2.6
Third Grade	5%	36%	18%	32%	9%	1.8	2.5
Fourth Grade	5%	38%	19%	38%		1.6	2.9
Fifth Grade	15%	35%	30%	20%		1.4	2.3

Question Three: Does coaching have an effect on the implementation process?

Question Four: Does the implementation of the elements of language arts training have a positive effect on the achievement of student in language arts?

The SEM Model

The Structural Equation Model (SEM) in Exhibit 5.1 is composed of boxes, circles or ovals, and arrows. The boxes indicate a measured variable and the circles or ovals indicate latent variables. A measured variable is the result of observable human behavior. A latent variable is something that cannot be observed directly but must be inferred from the results of observed behaviors. For

example, intelligence may be said to be a latent variable and the scores on various intelligence tests are observed variables.

The arrows in the model indicate the direction of causality. The direction of causality flows from the latent variables such as intelligence to the observed variables such as the results of intelligence test. This is the direction of causality because it is assumed that a person's innate intelligence causes the results of intelligence tests, that is, a person with high intelligence would receive high scores and a person with low intelligence would receive low scores on an intelligence test.

Variables in the Model
The following variables were included in the SEM model that was used to test the extent to which the training model was effective. All the measured variables are shown in mostly in lowercase letters and the latent variables are all in capital letters.

ACHIEVEMENT — This is a latent variable based on the level of achievement of students in language arts which includes writing and reading. The associated measures are Goals Performance Assessment Reading Scores and TAP writing scores.

- **Tapwrite** – a measured variable based on the Targeted Assessment Process (TAP) writing assessment given at the end of the school year to all students from 1st to 5^{th} grade. TAP is a writing assessment that uses a single writing prompt administered three times a year, once near the beginning, once in the middle and once at the end of the school year. The TAP is scored using a four-point rubric for kindergarten through third grade and a six-point rubric for student writing from fourth through twelfth grade. The 4^{th} -12^{th} grade four point rubric was modified so that absolute size difference did not affect the relative size of the covariance with the 1^{st} through 3^{rd} grade rubric scores. The score for each classroom is a mean average of all students' TAP writing scores in each classroom.

- **Goalread** – This is a measured variable based on the Goals Performance Assessment. This instrument contains a ten constructed response questions. The questions are based on two text passages, one of which is narrative and the other informational. The score for each classroom is a mean average of all students' Goals Performance Assessment reading scores.

- **Training** – This is a measured variable that is based on data collected by the training department on the number of hours that teachers have spent in training courses in the district.

IMPLEMENTATION — This is a latent variable generated based on the measured variables Wrtprcss, Wrtknow, and Readimp.

- **Wrtprcss** – This is a measured variable that is based on the Implementation Survey in the Writing Workshop part of the survey. The variable is a scale composed of questions 1 through 17.

- **Wrtknow** – This is a measured variable based on the Implementation Survey in the Writing Workshop part of the survey. The variable is a scale composed of questions 18 through 27.

- **Readimp** – This is a measured variable based on the Implementation Survey including all of the Reading Workshop section of the survey.

COACHING — This is a latent variable describing the level of support by district personnel that provide in-class support and feedback to teachers in language arts. The associated measured variables are Coachfrq, Litexpr, and Wrtknow.

- **Coachfrq** – This is a measured variable based on the responses to question 1 of the Language Arts Literacy Survey, which was collected from each teacher included in the analysis.

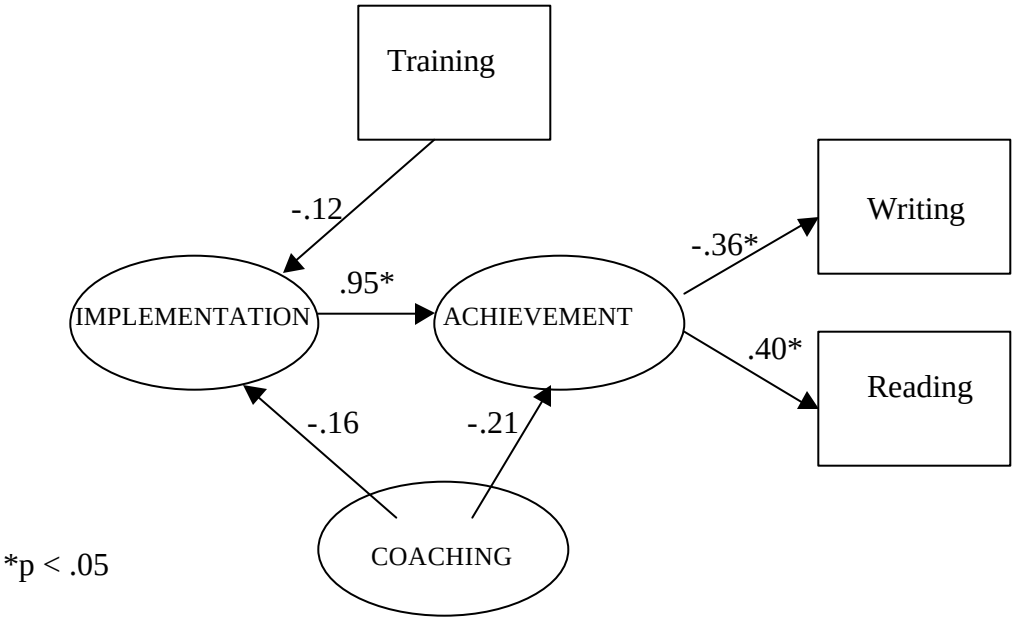
- **Litexpr** - This is a measured variable based on the responses to question 2 of the Language Arts Literacy Survey, which was collected from each teacher included in the analysis. It is part of the latent variable, COACHING.

- **Howused** - This is a measured variable based on the responses to question 3 of the Language Arts Literacy Survey, which was collected from each teacher included in the analysis. It is part of the latent variable, COACHING.

Statistical Analysis of the SEM Model

In the analysis, the model is statistically fit to determine whether the organization of variance and covariance in the model fits that which is found in the data that have been collected to represent the model. In other words, using the example of intelligence and related test scores, does the variation and covariation resulting from the test score relationships support the relationship between the test scores and the latent variable of intelligence? A Chi-square statistic provides a measure of fit. In the case of the training model which is the subject of this study, the Chi-square for the comparison between the model and the data was 27.006, with 25 degrees of freedom and a probability of .356. The alpha level for the statistical test was set at $p<.05$. Hence, because chi square is greater than .05, this statistic suggests that the model is a satisfactory fit to the covariance occurring in the data, and that the model accurately explains how the model of training used to explain what is happening in the data. The path coefficients shown in the model are generated as the program develops a simultaneous relationship of the measured and latent variables. The only significant relationships are between IMPLEMENTATION and ACHIEVEMENT, ACHIEVEMENT and Writing, and ACHIEVEMENT and Reading.

Exhibit 5.1: Structural Equation Model of Literacy Program Implementation



The strongest relationship is between IMPLEMENTATION and ACHIEVEMENT and there is a negative relation between ACHIEVEMENT and Writing. The other path coefficients essentially show that there are no other statistically significant relationships other than the three mentioned above. For example, there does not appear to be a significant path relationship between COACHING and IMPLEMENTATION nor between COACHING and ACHIEVEMENT.

The path diagram above shows only the primary components of the analysis described in terms of the standardized regression weights. Table 5.6 has been provided to show more complete statistics derived during the analysis of the complete Literacy Implementation model including all of the measured variables that were not provided in the path diagram.

Table 5.6
Complete Statistics for the Literacy Program Implementation Model

Dependent Variable (variance accounted for in measured variables)	Causal Variable	Regression Estimate	Standard Regression Weights	Standard Error	Critical Region
IMPLEMENTATION	COACHING	-0.490	-0.165	0.366	-1.340
IMPLEMENTATION	TRAINING	-0.203	-0.115	0.182	-1.115
ACHIEVEMENT	COACHING	-0.063	-0.206	0.061	-1.0 27
ACHIEVEMENT	IMPLEMENTATION	0.097	0.945*	0.026	3.679
litexper (.53)	COACHING	1.374	0.730	0.268	5.121
howused (.43)	COACHING	1.087	0.653	0.213	5.104
coachfrq (.48)	COACHING	1.000	0.695		
wrtknow (.62)	IMPLEMENTATION	1.103	0.787	0.186	5.943
wrtprcss (.64)	IMPLEMENTATION	1.000	0.799		
goalread (.16)	ACHIEVEMENT	1.000	0.403*		
readimp (.18)	IMPLEMENTATION	0.682	0.420	0.171	3.980
tapwrite (.13)	ACHIEVEMENT	-0.931	-0.362*	0.333	-2.794

* $p < .05$

Covariate Analysis

The covariate analysis used the 1998-99 Goals Performance Assessment Reading scores and Elementary School Proficiency Assessment (ESPA) for statistical control in the comparison of the 1999-00 results. It was felt that both the Goals and ESPA could be used because both tests include considerable reading and writing and would therefore reflect a more generalized language arts achievement. Average reading scores were compared for each teacher. This comparison was conducted separately for each grade.

In the first step in this statistical test, it was determined that scores differed significantly in each grade, i.e., one or more teachers had significantly different higher or lower scores than other teachers even when covariance controls had been applied for prior achievement. First grade was not included in the analysis because there were no scores from the previous year with which to control for prior achievement. After identifying statistical differences among the teachers in each grade from 2^{nd} to 5^{th}, a post-hoc comparison was used to identify the class averages at each grade that were responsible for the statistically significant results.

A fixed effects analysis of covariance was used to test the Goals Performance Assessment reading scores of grades 2, 3, and 5 and the language arts ESPA scores of grade 4 for difference ($p<.001$) across teachers. When a significant difference was found, a test of multiple comparisons was used to identify those classrooms with scores significantly higher or lower than the expected mean for the entire group of scores. Each grade was tested separately. The covariate used was the previous years achievement test scores. Once the classes with the higher and lower than average scores were identified, a chi-square test was conducted to determine whether the classes were consistently associated with high and low levels of implementation. The implementation measurements used were the same data that was used in the path analysis

explained previously. The chi-square was not significant which suggests that the level of implementation, although related, does not appear to be responsible for significantly higher or lower scores on the language arts achievement tests. The results by grade are shown below. Tables 5.7 through 5.10 show the analysis of covariance for each of the grades indicating the level of significance identified in the average teacher language arts scores referred to above.

Table 5.7
ANCOVA Across Teachers' Language Arts Literacy Score Averages – Grade 2

Source	Sum of Squares	Df	Mean Square	F	Sig.
Corrected Model	490904.556	28	16363.485	6.958	.000
Intercept	254310.095	1	254310.095	108.140	.000
Reading Score	194570.403	1	194570.403	82.737	.000
Homeroom	205045.333	27	7070.529	3.007	.000
Error	1044146.168	407	2351.681		
Total	159946477.000	436			
Corrected Total	1535050.724	435			

Table 5.8
ANCOVA Across Teachers' Language Arts Literacy Score Averages – Grade 3

	Sum of Squares	Df	Mean Square	F	Sig.
Corrected Model	241417.732	28	8622.062	5.415	.000
Intercept	454674.587	1	454674.587	285.558	.000
Reading Score	137724.793	1	137724.793	86.489	.000
Homeroom	93156.436	27	3450.238	2.167	.001
Error	648037.687	407	1592.230		
Total	153806629.000	436			
Corrected Total	88945.420	435			

Table 5.9
ANCOVA Across Teachers' Language Arts Literacy Score Averages – Grade 4

Source	Sum of Squares	Df	Mean Square	F	Sig.
Corrected Model	110131.441	311	3552.627	8.342	.000
Intercept	1750.518	1	1750.518	4.110	.043
Reading Score	58558.668	1	58558.668	137.501	.000
Homeroom	42945.909	30	1431.530	3.361	.000
Error	129041.11	303	425.878		
Total	11206145.000	335			
Corrected Total	2391723.555	334			

Table 5.10
ANCOVA Across Teachers' Language Arts Literacy Score Averages – Grade 5

Source	Sum of Squares	Df	Mean Square	F	Sig.
Corrected Model	366958.350	29	12653.736	8.992	.000
Intercept	1551662.780	1	1551662.780	1102.702	.000
Reading Score	216421.883	1	216421.883	153.802	.000
Homeroom	111048.162	28	3966.006	2.818	.001
Error	481243.897	342	1407.146		
Total	173746138.000	372			
Corrected Total	848202.147	371			

In an attempt to better understand the level of the significance of the relationship between implementation and achievement shown in the SEM analysis, the following comparison was developed. This analysis was meant to understand whether teachers with significantly higher scores did in fact do a noticeably better job of implementation. The SEM analysis of the literacy program implementation showed a relationship between implementation and achievement. The analysis shown in Tables 5.7 through 5.10 was designed to determine how extensively observable the relationship was. Table 5.11 provides a comparison between the teachers that were identified as showing significantly high or low language arts scores with the level of implementation. The two levels of implementation were derived by dividing the Implementation scores in two groups. Teachers with higher scores were those above the average implementation score in reading and the lower scores were those below the average implementation score in reading. In the chart below there were a total of 21 teachers in the analysis. Of these 21, 12 showed significantly high language arts scores and 9 showed significantly low language arts scores. There were 13 of the 21 who had higher than average implementation scores and 8 with lower than average implementation scores.

Table 5.11
A Chi-Square Analysis of the Relation between Language Arts Achievement and Implementation

	Higher Than Average Implementation Scores	**Lower Than Average Implementation Scores**	**Totals in Language arts scores**
Significantly High Language Arts Scores	8	4	12
Significantly Low Language Arts Scores	5	4	9
Totals in implementation	13	8	21

Chi-square = .269, p = .673

The chi-square analysis results were not significant, suggesting that the relationship shown in the SEM model is relatively slight. It appears that there might be some slight difference, in that eight of the high scoring teachers were in the upper half of the implementation range.

Educational Value-Added Assessment System (EVAAS)

EVASS is a process that measures the influence that systems, schools and teachers have on the rate of academic growth for populations of students. In this evaluation study, EVAAS is being used to identify those teachers who have been effective over the past three years and compare the effective versus less effective teachers with their students' responses on the Literacy Implementation survey and the teachers' responses to the Literacy Coaching Experiences Survey. To accomplish this analysis, the EVAAS uses a statistical mixed-model methodology and student scale scores from norm-referenced achievement tests and any other measure of academic growth that is strongly related to the curriculum. EVAAS's mixed model is capable of overcoming aspects of the assessment that would bias results when the factors can be identified that have a biasing effect on results. For example, effects attributable to individual systems and schools appear unrelated to socio-economic indicators such as number of free- or reduced-price lunch students, racial composition of the student body, or urbanicity.

EVAAS uses scale scores to establish where a child is academically and to determine how much progress that child makes in each school year. EVAAS concentrates on gains that provide information on educational effects. The model provides an analysis that concentrates on these gains instead of artifacts related to normative interpretations or the size of particular students scores.

EVAAS proved a much more conservative test than the covariance analysis described above. In addition, the data were only sufficiently reliable in grades three and five to meet the initial criterion of the EVAAS analytic model. This model identified eight teachers that showed either significantly higher or lower scores than the general population of teachers as a whole ($p < .05$). All eight of those identified by EVAAS were also among the 21 teachers identified by the covariance analysis described above (see Table 5.11). Since EVAAS represented a more restrictive but unfortunately limited test of the achievement, it was decided to defer to the covariance analysis and not pursue any further analysis of the teachers identified as showing exceptionally high or low student achievement.

Successes and Challenges

Successes

The successes of this study relate not to the findings but to the efforts that people have made in relation to the data that were used and the simple fact that people have begun to think about data as relevant to what they do as teachers. Before this effort, few if any teachers felt that the results of district-wide assessments had any relevance to what they do in the classroom. This evaluation effort and the involvement required has helped to show some people that these kinds of data do have some relevance to what is occurring in the classroom.

Getting district staff to participate in data collection and data analysis was difficult. It is hoped that the effort spent and successive efforts at involvement will be more meaningful for teachers especially if results of future analyses show achievement levels responding to the successive training efforts. There is continuous effort to raise teachers' awareness of data, data use, and data quality. The increase focus on data use has helped teacher to begin thinking about how data can relate to programs instead of only individual children. As they develop this concept, they have also begun to enhance their own models of how the assessment and evaluation can function in the context of their school environments.

Process Challenges

Although some progress has been made in getting administrators and teachers to understand the relationship between evaluation and programs, the progress has not been complete. It has also been in relation to relatively limited aspects of education. So the challenge is to help teachers and administrators generalize their understanding. This generalizing could possibly improve if they become involved in developing models for district use of evaluations, i.e., thinking about ways they themselves could find the results of evaluations useful for their own programs. But the process will never become completely part of their normal operations until it becomes apparent to them that assessment and evaluation make program improvement more efficient and ultimately results in a better achievement outcome.

Data Challenges

The following list illustrates some of the problems encountered with the data that will need to be considered when interpreting the results of the analyses in this report:

- Reliability The assessment instruments in the study, Goals and TAP, showed low reliability - depending on the grade. The internal consistency reliability was between 0.63 and 0.71. The TAP Reading assessment also showed low internal consistency reliabilities of between 0.58 and 0.70. The TAP Writing assessment is a single-item test, so internal consistency reliability for this test cannot be calculated. Pearson Product Moment correlations between the 4th Grade TAP writing and the ESPA writing to a picture produced coefficients of 0.29 and 0.21 in writing to a poem. Inter-rater reliability is not calculated for the scoring of either the Goals Performance Assessment, TAP Reading, or TAP Writing although an analysis of a sample of responses was examined and the correlations were below 0.60. Generalizability of the TAP writing results is never considered even though it would be possible to conduct the analysis without much difficulty.

- Validity - There has been no internal validation by the district of the Goals Performance Assessment or TAP to determine whether it is measuring important aspects of the curriculum.

- Teachers' Response Rates- There was considerable reluctance among teachers to complete the one-page surveys, with most of the teachers having to be sent the forms at least three times and principals being asked to intervene in a number of instances. Only 57 percent of the surveys were returned.

- Training Data - Training data were not kept in a manner that one could readily determine the number of hours that each teacher spent on literacy training. The training data collected were generated by examining a number of different documents, some of which appeared to be contradictory when compared to the data that were collected on the Literacy Coaching Experiences Survey.

- Misclassifications - Student files often contained misclassifications and incorrect or missing information. For example, many special needs children are not so labeled in the computer file and their test scores are then included with the non-classified regular students.

- Incomplete data – District files are eliminated annually, so that a number of longitudinal comparisons were not available. This made it impossible to complete some the tasks that were originally planned. For example, the Educational Value-Added Assessment System (EVASS) procedure was not possible because yearly teacher identifications were not available.

- Interpretation Differences – There was considerable variation in the way that individuals interpreted concepts associated with the language arts program. For example, teachers held different ideas about the conceptualization and implementation methods associated with basic constructs of the evaluation such as coaching.

  Additionally, the Language Arts Literacy Survey was developed because of differences in interpretation of the classroom observations, which were originally planned to be used. These observations were conducted by district and school administrators in every classroom in the

district. Nothing was done, however, to ensure that the observations of the observers was consistent. The district decided not to use these data because of the wide variety of interpretations used. The Implementation Survey was completed by one person across all of the schools. It would have been much better to use direct observations, but the inconsistency of interpretation was deemed too great to be useful.

Conclusions

Question 1: Is training in language arts being implemented as designed?

The results show that program implementation is relatively low, as reported by students. The 1999-2000 school year was the first year of the implementation, consequently many aspects of the program may not have been functioning at the level planned. It was expected that the writing workshop would show a higher level of implementation than reading workshop since more attention was paid to writing workshop during the 1999-2000 school year. However, this is not what results showed. This may have been a function of length of the protocols used – the writing workshop protocol had more questions to answer than the reading workshop protocol. This is because the writing workshop includes more components than the reading workshop.

The results showed more consistency in the implementation of the writing workshop than the reading workshop – this may have been as mentioned above the result of a greater focus on the development of the writing workshop during the 1999-00 school year and curricular efforts in all schools contained more emphasis on writing than on reading.

Reading training appears to be very uneven. This may be a result of less attention being paid to reading since it was not the focus of the language arts program in any of the schools during the 1999-00 school year.

Question 2: Is coaching in language arts working as designed?

For the 1999-2000 school year, Cook, Woodland and Cedarbrook, seem to pay the most attention to the different aspects of the coaching program given the results across Coaching Frequency, Coaching Experience and Coaching Use. Barlow, Jefferson, and Stillman appear to pay the least attention to the different aspects of the coaching program. The schools with low frequency of coaching also appear to experience fewer experiences with and uses of coaching.

In 1999-2000, coaching does not appear to be as extensively used as it could be, particularly in the schools with the lowest levels of coaching frequency. Coaching appears to be most frequent and well used in the schools associated with the Whole School Reform program coordinated by America's Choice. The non-America's Choice schools — Jefferson, Stillman, and Washington, which do not have the level of emphasis on coaching in their program as schools involved in America's Choice — were by and large lower in their levels of frequency, use and experience of coaching. Hence, the schools with the greatest emphasis on coaching in their programs do appear to be experiencing the greatest frequency, use and experience with coaching. None of the schools, however, appear to be using coaching to the extent that it could be used. Perhaps this could be explained by the fact that both America's Choice is relatively new and there might be

some confusion about what to expect and how to proceed although the coaching program has been functioning for over two years in much of the district.

Question 3: Does coaching have an effect on the implementation process?

The results of the structural linear equation model suggest that during 1999-00, coaching did not have an effect on either the implementation of language arts or language arts achievement. As shown in Table 5.6 and Exhibit 5.1, the coefficient between coaching and implementation was not statistically significant (-.16) and between coaching and achievement was also statistically not significant (-.21). In addition, training, as measured, does not appear to have had an effect on implementation. It is difficult to tell whether this finding is a result of what occurred in the coaching or the problems with measuring the coaching. This question will not be satisfactorily answered until the problems with the data are addressed.

Question 4: Does the implementation of the elements of language arts training have a positive effect on the achievement of student in language arts?

Based on the results to the SEM analysis, it appears that coaching does not have an effect on implementation or on achievement. There was however, a significant path coefficient between implementation and achievement. To examine this relationship further, and determine to what extent higher levels of implementation result in significantly higher language arts achievement, the covariance analysis and EVAAS were performed. The level of implementation, however, did have an effect on student achievement. The standardized path coefficient between implementation and achievement was .95.

The results of the SEM model suggest that TAP writing scores were negatively correlated with achievement. This result may be explained by the one-item nature of the writing score. On the other-hand, the reading scores were positively correlated. A Pearson Correlation Coefficient was conducted on these two sets of scores to get another perspective on the relationship and it showed -.15 which is significant at $p<.5$ but is a very weak correlation between the Goals reading and TAP writing. This low correlation is reflected in the high amount of residual error associated with each variable shown in the SEM results.

Subsequent investigation with the implementation results suggests that the relationship between implementation and achievement is not enough to produce significantly higher language arts achievement results.

District Response to Study Conclusions

Based on our investigation of the four research questions, it appears that none of the aspects of the language arts program are having the effect for which they were designed. It is probably too early to expect to find effects on student achievement. This is problematic, however, because the lack of effects may also be the result of the number of data errors discussed earlier. If the data are to be accepted, these non-significant results suggest that the program had no effect during the 1999-2000 school year. To increase confidence in future evaluation results, it would be important for the district to focus on collecting data that are more valid. For example, the use of

a larger sample of writing responses may increase the validity of results by better sampling the domain of writing on which the statistical relations depend. In subsequent studies, with more reliable and valid data, the district should look more closely at the coaching program. It might also help if formative evaluations were conducted to determine whether aspects of the language arts program, such as coaching, are being conducted as designed. Triangulation across different sources of data on specific, important issues could be used to support the level of accuracy of the data collected.

Actions Taken

The basic lesson of these data for the district is that there is considerable need to clean up much of the data collection process and improve the understanding of the various aspects of the data that are being produced. Towards this end, the district is in the process of developing a data warehouse to provide better data for the kinds of analyses that were attempted in this evaluation, which will be operational in by next year. At this time, the data will be re-evaluated to determine if the evaluation should be duplicated in the future. As for the near future, the district decided to continue the program with few if any modifications. It was also felt that it was too soon in the development of the language arts program and the data on which the analysis needed to be improved before a replication of the study could be accomplished and perhaps compared to the current study.